Rapid, Machine-Learned Resource Allocation: Application to High-redshift GRB Follow-up

A. N. Morgan[1,*], James Long[2], Joseph W. Richards[1,2], Tamara Broderick[2],
Nathaniel R. Butler[1,3], and Joshua S. Bloom[1]

ABSTRACT

As the number of observed Gamma-Ray Bursts (GRBs) continues to grow, follow-up resources need to be used more efficiently in order to maximize science output from limited telescope time. As such, it is becoming increasingly important to rapidly identify bursts of interest as soon as possible after the event, before the afterglows fade beyond detectability. Studying the most distant (highest redshift) events, for instance, remains a primary goal for many in the field. Here we present our Random forest Automated Triage Estimator for GRB redshifts (`RATE GRB-`z) for rapid identification of high-redshift candidates using early-time metrics from the three telescopes onboard *Swift*. While the basic `RATE` methodology is generalizable to a number of resource allocation problems, here we demonstrate its utility for telescope-constrained follow-up efforts with the primary goal to identify and study high-z GRBs. For each new GRB, `RATE GRB-`z provides a recommendation—based on the available telescope time—of whether the event warrants additional follow-up resources. We train `RATE GRB-`z using a set consisting of 135 *Swift* bursts with known redshifts, only 18 of which are $z > 4$. Cross-validated performance metrics on this training data suggest that ∼56% of high-z bursts can be captured from following up the top 20% of the ranked candidates, and ∼84% of high-z bursts are identified after following up the top ∼40% of candidates. We further use the method to rank 200+ *Swift* bursts with unknown redshifts according to their likelihood of being high-z.

Subject headings: Gamma-ray burst: general – Methods: data analysis – Methods: statistical

[1]Department of Astronomy, University of California, Berkeley, CA 94720-3411, USA.

[2]Department of Statistics, University of California, Berkeley, CA 94720-3860, USA

[3]Department of Physics, Arizona State University, Tempe, AZ 85287, USA

[*]amorgan@astro.berkeley.edu

1. Introduction

As the most luminous electromagnetic explosions, gamma-ray bursts (GRBs) offer a unique probe into the distant universe—but only if their rapidly fading afterglows are observed before dimming beyond detectability (e.g., Wijers et al. 1998; Miralda-Escude 1998; Lamb & Reichart 2000; Kawai 2008; McQuinn et al. 2008). Since the launch of the *Swift* satellite in November 2004 (Gehrels et al. 2004), more than 170 long duration *Swift* gamma-ray bursts have had measured redshifts, but only a handful fall into the highest redshift range that allow for the probing of the earliest ages of the universe, up to less than a billion years after the Big Bang (Fig. 1). With a limited budget of large-aperture telescope time accessible for deep follow-up, it is becoming increasingly important to rapidly identify these GRBs of interest in order to capture the most interesting events without spending available resources on more mundane events.

Along with quasars (e.g., Mortlock et al. 2011) and NIR-dropout lyman-break galaxies (e.g., Bouwens et al. 2010, 2011), GRBs have been established as among the most distant objects detectable in the universe, with a spectroscopically confirmed event at $z = 8.2$ (GRB 090423; Tanvir et al. 2009; Salvaterra et al. 2009) and a photometric candidate at $z \sim 9.4$ (GRB 090429B; Cucchiara et al. 2011b). Such observations can provide valuable constraints on star formation in the early universe, illuminate the locations and properties of some of the earliest galaxies and stars, and probe the epoch of reionization. (e.g., Tanvir & Jakobsson 2007, and references therein). Further, the relatively simple spectra of GRB afterglows compared to other cosmic lighthouses makes it easier to both identify their redshifts and extract useful spectral features such as neutral hydrogen absorption signatures for the study of cosmic reionization. (e.g., Miralda-Escude 1998; Barkana & Loeb 2004; Totani et al. 2006; McQuinn et al. 2008). However, such benefits can only be realized if spectra are obtained with large-aperture telescopes before the afterglow fades beyond the level required to obtain a useful signal, typically within a day after the GRB.

As such, there has been a long-standing effort to extract a measure of a GRB's redshift from its early time, high-energy signal, with a primary goal of the rapid identification of high-z candidates. This might appear in principle to be a straightforward exercise; for instance, distant GRBs should on average appear fainter and longer-duration than nearby events due to distance and cosmological time dilation, respectively. In practice, however, the large intrinsic diversity of GRBs, as well as thresholding effects, confounds the straightforward use of early-time observations in divulging redshift and other important properties. While much effort has gone into tightening the correlations between high-energy properties in order to homogenize the sample for use as a luminosity (and hence distance/redshift) predictor (e.g., Amati et al. 2002; Ghirlanda et al. 2004; Firmani et al. 2006; Schaefer 2007), there has

been significant debate as to whether some of these relations are actually due to thresholding effects specific to the detectors rather than intrinsic physical properties of the GRBs (e.g., Friedman & Bloom 2005; Butler et al. 2007, 2009). Regardless, whether or not these inferred relationships are actually physical or simply detector effects would not affect their utility as a *detector-specific* parameter prediction tool. By restricting ourselves to *Swift* events only, we avoid the uncertainty of whether certain correlations remain when using different detectors.

With this in mind, we set out to search for indications of high-redshift GRBs in the rich, mostly homogeneous dataset provided by 6+ years of GRB observations by the three telescopes onboard *Swift* (BAT; Barthelmy et al. 2005, XRT; Burrows et al. 2005, UVOT; Roming et al. 2005). Past studies exploring high-z indicators have used hard cuts on certain features such as UVOT afterglow detection, burst duration, and inferred hydrogen column density (e.g., Grupe et al. 2007; vanden Berk et al. 2008; Ukwatta et al. 2009), regression on such features (Koen 2009, 2010), and combinations of potential GRB luminosity indicators (Xiao & Schaefer 2009, 2011). In this work, we take a different approach by utilizing supervised machine learning algorithms, specifically Random Forest classification, to make follow-up recommendations for each event automatically and in real time. Particular attention is paid to careful treatment of performance evaluation by using cross-validation (§4), a robust methodology to guard against over-fitting and the circular practice of testing hypotheses using the same data that suggested (and constrained) them.

The primary driving force of this study is simple: *given limited follow-up time available on telescopes, we want to maximize the time spent on high-z GRBs*[1]. To this end, we provide a deliverable metric, explained in §3.2, to assist in the decision making process on whether to follow up a new GRB. Real-time distribution of this metric is available for each new *Swift* trigger via website[2] and RSS feed[3].

The structure of this paper is as follows: in §2 we outline the collation of the data, and describe the particular GRB features utilized in redshift classification. In §3, the Random Forest algorithm is detailed, along with some specific challenges posed by this particular data set. Performance metrics of the classifiers are presented in §4, and in §5 we discuss the results of testing the classifiers on additional GRBs, both with and without known redshifts. Finally, our conclusions are given in §6.

---

[1]For the purposes of this study, "high-redshift" corresponds to all $z > 4.0$: a compromise between only keeping the most interesting events and having enough data to train on. However, we have explored performance of different redshift cuts; see §4.3.

[2]http://rate.grbz.info/
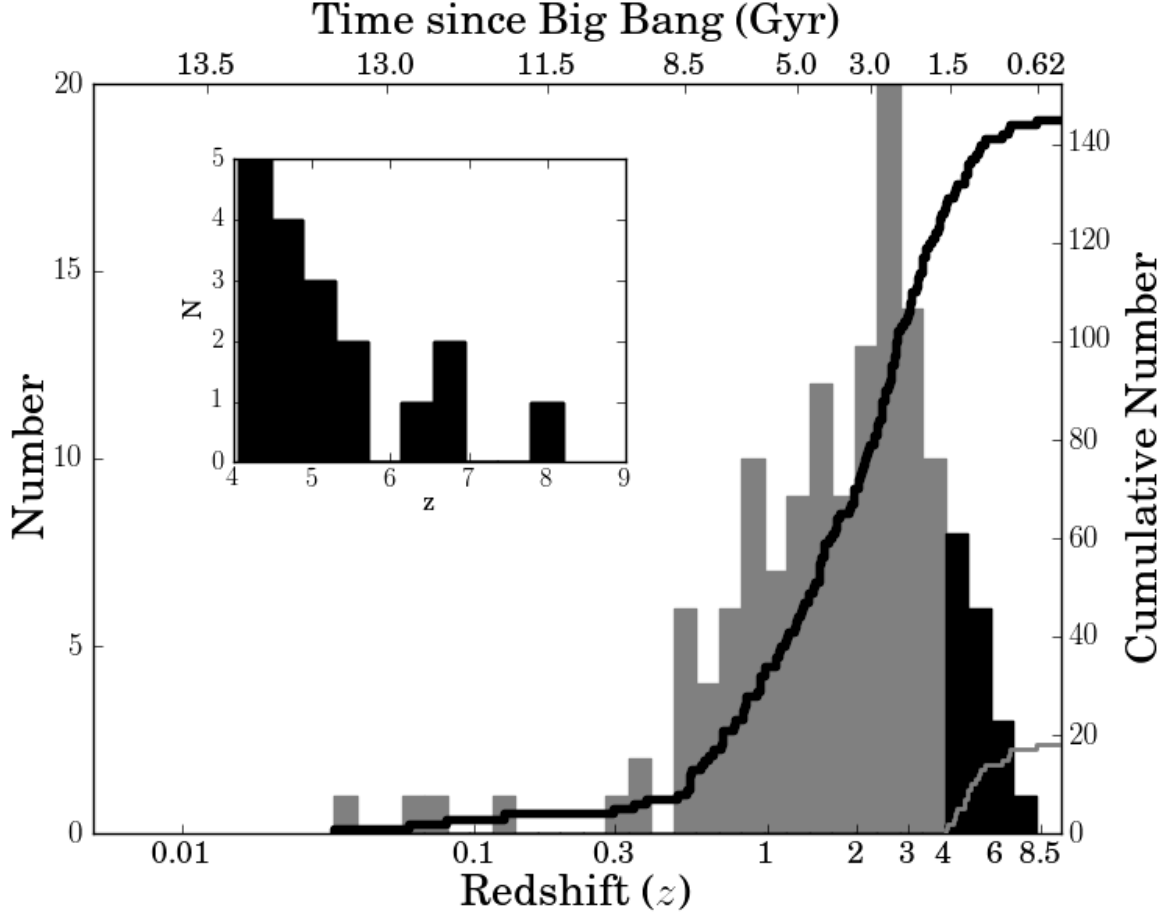
[3]http://rate.grbz.info/rss.xml

Fig. 1.— Redshift distribution of the 135 long-duration *Swift* GRBs in our sample (Table 2). For the purposes of this study, "high" redshift is defined as those bursts with redshifts larger than $z = 4$, which corresponds to approximately 1-σ above the mean of the distribution. In our sample, 18 bursts fall into this category (black, and in inset). In determining age since the Big Bang, we assume a cosmology with $h = 0.71$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$. Solid lines show the cumulative number of GRBs as a function of redshift for high-z bursts (grey) and all bursts (black).

2. Data Collection

The *Swift* BAT constantly monitors 1.4 steradians on the sky over the energy range $15 - 150$ keV. GRB triggering can occur either by a detection of a large gamma-ray rate increase in the BAT detectors ("rate trigger"), or a fainter, long-duration event recovered after on-board source reconstruction reveals a new significant source ("image trigger"). A rough ($\sim$

3 arcmin) position is determined, and if there are no overriding observing constraints, the spacecraft slews to allow the XRT and UVOT to begin observations, typically between 1 and 2 minutes after the trigger. The XRT observes between the energy range of $0.2 - 10$ keV and detects nearly all of the GRBs it can observe rapidly enough, providing positional accuracies of $2 - 5$ arcseconds within minutes. The UVOT is a 30cm aperture telescope that can observe in the range of $170 - 650$ nm. Due to the relatively blue response of this telescope, it cannot detect highly reddened sources due to either dusty environments or (more relevant to this analysis) high-redshift origins.

At each stage in the data collection process, information is sent to astronomers on the ground via the Gamma-ray bursts Coordinates Network (GCN[4]) providing rapid early-time metrics. The more detailed full data are sent to the ground in ~ 90 minute intervals starting between roughly $1 - 2$ hours after the burst. For our dataset, we have collected data after various levels of processing directly from GCN notices, online tables[5] and automated pipelines (Butler & Kocevski 2007; Butler et al. 2007) that process and refine the data into more useful metrics. Tens of attributes and their estimated uncertainties (when available) are parsed from the various sources and collated into a common format.

In order to evaluate our full dataset in an unbiased way, we restricted ourselves to using features which have been generated for all possible[6] past events and are automatically generated for future events. This is the primary reason we do not include potentially useful features such as relative spectral lag (e.g., Ukwatta et al. 2010, 2011, and references therein) which has been utilized as a redshift indicator with smaller and pre-*Swift* datasets (Murakami et al. 2003; Band et al. 2004; Zhang et al. 2006; Schaefer 2007) but requires a larger spectral coverage than *Swift* alone can provide. However, our technique is easily extendable to include additional useful features should they be homogeneously determined for past GRBs and automatically available in real-time for new events, and therefore we strongly encourage the automated distribution of any such data products.

Because the addition of too many features causes a decrease in classifier performance (see §4.2), a total of 12 features were kept for our final classifier (Table 1), 10 of which were derived from BAT gamma-ray measurements, one from XRT observations, and one from UVOT observations. Of the 10 BAT features, 4 were parsed directly from GCN Notices, the

---

[4]http://gcn.gsfc.nasa.gov/

[5]http://swift.gsfc.nasa.gov/docs/swift/archive/grb_table.html/

[6]Even with the restriction of observation by all 3 *Swift* telescopes, certain features derived from model fits are nonetheless incalculable for certain GRBs from the available data. See §3.1.1 for how our algorithm treats missing values.

most rapidly available (and thus unrefined) source of information on GRBs[7]. The parameter t_{BAT} is a rough measurement of the duration of the BAT trigger event and thus a lower limit on the total duration of the GRB. The binary feature of whether or not the event was a rate trigger is an indicator of the signal-to-noise of an event, for only the brighter events are detected as rate triggers, while those on the threshold of detection are image triggers. The final two GCN features are also rough indicators of brightness: σ_{BAT} is the significance (in sigma) of the detected source in the on-board reconstruction of the BAT image, and $R_{peak,BAT}$ is the peak count rate observed during the duration of the event.

Five higher-level BAT-derived attributes were pulled from online tables automatically updated by the pipeline described in Butler et al. (2007). The feature α is the power-law index before the peak of the Band-function fit to the gamma-ray spectrum (typically clustered around -1). Another parameter in the Band-function fit, E_{peak}, is the energy at which most of the photons are emitted. The fluence, S, is the total gamma-ray flux (15–350 keV) integrated over the duration of the burst. S/N_{max} is simply the maximum signal-to-noise achieved over the duration of the light curve. Finally, T_{90} is a measure of the burst duration, defined to be the time interval over which the middle 90% of the total background-subtracted flux is emitted.

One additional "metafeature" is derived from the BAT data. In principle, if we knew in detail the intrinsic distributions of GRB observables (fluence, hardness, duration; see Butler et al. 2007) as a function of redshift, measurements of these observables for a new event could be used to directly evaluate the expected redshift. A detailed fitting of the intrinsic distributions for Swift is presented in Butler et al. (2010), and we use the parametrized intrinsic distributions there to calculate the posterior probability redshift distributions for each GRB in our sample (see, e.g., Figure 8 in Butler et al. 2010). Here, we further condense this distribution into one useful feature: $P_{z>4}$, the fraction of posterior probability at $z > 4$.

Finally, two features are extracted from data taken by the two narrow-field instruments onboard *Swift*, one each from the XRT and UVOT. The feature $N_{\mathrm{H,pc}}$ is the excess neutral hydrogen column (above the galactic value) inferred from the XRT PC (Photon-counting mode) data, obtained from the Butler & Kocevski (2007) pipeline. The last feature is simply a binary measure of whether or not the GRB afterglow was detected by the UVOT.

While most of these features have associated uncertainties, the proper treatment of un-

---

[7]For 14 events in our test set, the SWIFT_BAT_POSITION notice was not available on the online repository, primarily due to satellite downlink problems at the time of discovery. For these events, the relevant parameters were extracted directly from the *Swift* TDRSS database (http://heasarc.nasa.gov/W3Browse/all/swifttdrss.html).

certainties in attributes is an area of ongoing research in machine learning (e.g. Carroll et al. 2006). Some methods call for the uncertainties to be treated as attributes in and of themselves, but we found that the addition of these relatively weak features were actually detrimental for our small dataset (see, e.g., Fig. 6). We also considered an approach by which features with large uncertainties were considered poor measurements and were instead marked as missing values. However, this had a negligible effect on our final classifier performance, so for simplicity we treat all values as precisely known.

We collated data on all *Swift* GRBs with rapidly available BAT data up to and including GRB 100621A - 471 in total. Specifically, this excludes bursts which were not identified in real-time due to the event being below the standard triggering threshold or occurring while the satellite was slewing to a new location. Of these, 39 are short GRBs (defined for the purposes of this study to be those with $T_{90} < 2.0$ s[8]), which are believed to arise from a different physical process and are thus removed from the sample. For further uniformity in the sample, bursts without rapid (< 1 hour) XRT/UVOT follow-up are also removed, leaving 347 events[9]. Of the remaining long bursts in our sample, 135 had reliable redshifts (Table 2) and were thus included in our training data set (Table 3). The additional 212 long bursts without secure redshift determinations are explored further in §5.1. Exploratory data analysis shows preliminary indications of which of these features will be most useful for classification. Figure 2 shows several 2D slices of the feature space, with the high-z bursts highlighted.

---

[8] T_{90} alone is not a strong enough discriminator to definitively assign a particular GRB to one class or another ("short" versus "long"; see Levesque et al. 2010 for discussion). In this study, we will accept the few errant bursts from the "short" class included in our sample as additional noise in our method.

[9] The reason for this missing data is almost always due to observing constraints from the GRB being too close to the Sun, Moon, or Earth at the time of discovery. Not removing these bursts would introduce a bias in the sample due to the fact that events without a rapid XRT position are far less likely to lead to an afterglow discovery, and hence, redshift determination. A total of 15 bursts with known-z were removed because of this.

Table 1.   List of Features Utilized

Feature	Type	Reference
BAT Rate Trigger?	BAT Prompt	GCN Notices
σ_{BAT}	BAT Prompt	GCN Notices
$R_{peak,BAT}$	BAT Prompt	GCN Notices
t_{BAT}	BAT Prompt	GCN Notices
UVOT Detection?	NFI Prompt	GCN Notices
$N_{\mathrm{H,pc}}$	Processed	Butler & Kocevski (2007)
α	Processed	Butler et al. (2007)
E_{peak}	Processed	Butler et al. (2007)
S	Processed	Butler et al. (2007)
S/N_{max}	Processed	Butler et al. (2007)
T_{90}	Processed	Butler et al. (2007)
$P_{z>4}$	Processed	Butler et al. (2010)

Table 2.   Training Data Redshifts

GRB	$\widehat{\mathcal{Q}}_{train}$	z	References
050223	4.30e-01	0.5915	Berger & Shin 2006
050315	3.57e-01	1.949	Kelson & Berger 2005
050318	6.86e-01	1.44	Berger & Mulchaey 2005
050319	5.90e-01	3.2425	Fynbo et al. 2005a; Jakobsson et al. 2006c; Fynbo et al. 2009b
050416A	7.68e-01	0.6535	Cenko et al. 2005

Note. — Table 2 is published in its entirety in the electronic edition of The Astrophysical Journal. A portion is shown here for guidance regarding its form and content.

Table 3.   Training Data

GRB	α	E_{peak} (keV)	S (erg/cm^2)	S/N_{max}	$N_{H,pc}$ (10^{22} cm^{-2})	T_{90} (s)	σ_{BAT}	$R_{peak,BAT}$ (ct/s)	Rate trigger	t_{BAT} (s)	UVOT detect	$P_{z>4}$
050223	-1.74e+00	6.70e+01	8.75e-07	1.34e+01	-2.37e-01	1.74e+01	9.00e+00	7.26e+02	yes	8.19e+00	no	1.74e-01
050315	?	4.33e+01	4.32e-06	4.37e+01	9.60e-02	9.46e+01	8.00e+00	2.60e+02	yes	1.02e+00	no	9.27e-02
050318	-1.22e+00	5.01e+01	1.41e-06	4.90e+01	1.80e-02	3.10e+01	9.00e+00	2.05e+02	yes	5.12e-01	yes	6.29e-02
050319	-2.00e+00	4.47e+01	1.87e-06	1.82e+01	1.50e-02	1.54e+02	1.00e+01	2.63e+02	yes	1.02e+00	yes	1.48e-01
050416A	-7.24e-01	1.50e+01	3.40e-07	1.75e+01	2.34e-01	2.91e+00	1.10e+01	1.65e+02	yes	5.12e-01	yes	4.35e-03

Note. — Table 3 is published in its entirety in the electronic edition of The Astrophysical Journal. A portion is shown here for guidance regarding its form and content.
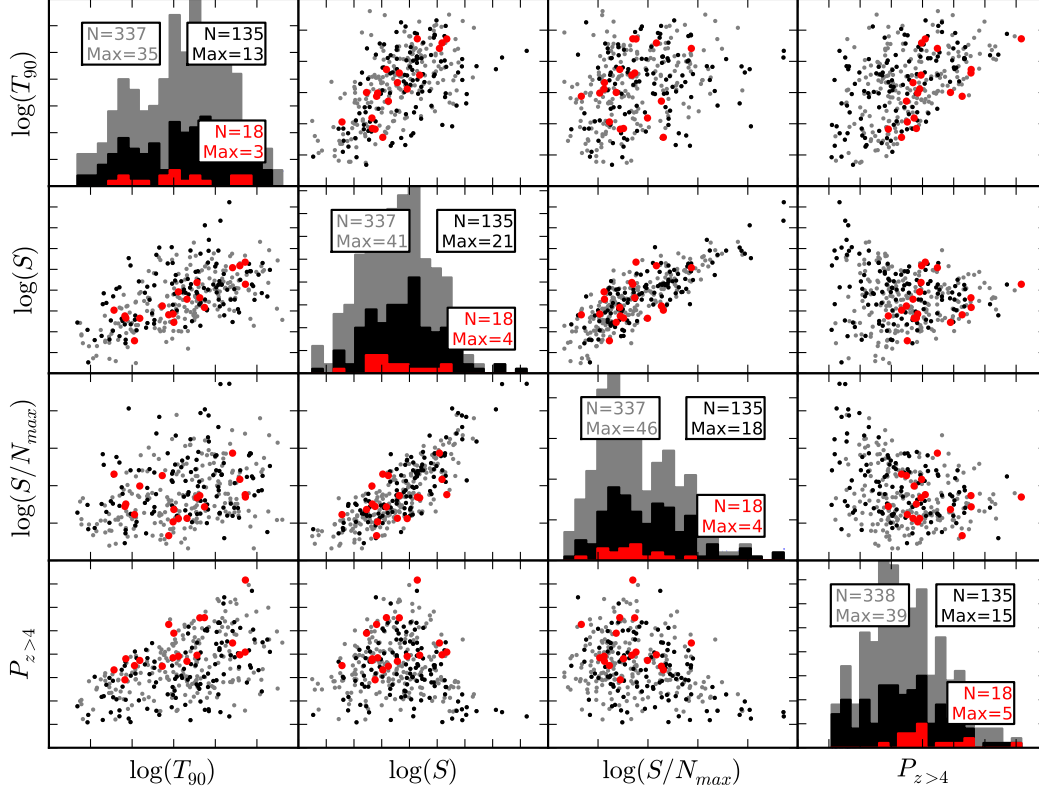
Fig. 2.— Plot of a selection of early-time *Swift* features (Table 1) against each other. The grey points show the full distribution of Swift GRBs. Bursts with known redshifts are black, and the 18 known events with redshifts greater than 4 are overplotted in red. In the histogram text boxes, N shows how many instances of that feature in total are shown (anything less than the full number of instances is due to the value of that feature being unknown for certain instances), and Max shows the maximum number of instances in any particular bin.

3.   Classification Methodology

The resource allocation approach we have taken here naturally manifests itself as a classification problem: deciding whether or not to follow up a new event is simply a two-class problem of "observe" or "do not observe," and the methodology presented here can be applied to any problem that can be broken up in this way. This was the primary motivation of

using classification instead of a regression or "pseudo-z" approach for this study. The primary disadvantage of classification for the particular problem of high-redshift identification is that all instances above and below the class division (chosen here to be $z = 4$) are treated equally; e.g., a burst with $z = 4.01$ has the same influence on our inference about "high" bursts as a burst with $z = 8^{10}$. However, classification has advantages over regression in that it is a conceptually much simpler problem, and most of the difficulties encountered due to the unbalanced, small dataset of interest here would only be aggravated by an extension to regression. Further, our approach capitalizes on the fact that one of our predictors (lack of UVOT detection) is itself a binary feature with an understood physical connection to redshift[11].

3.1. Random Forest classification

A supervised classification algorithm uses a set of training data of known class to estimate a function for assigning data points to classes based on their features. The statistics and machine learning communities have developed many classification algorithms, including Support Vector Machines (SVM), Naïve Bayes, Neural Networks, and Gaussian Mixture Models. We use Random Forest (RF Breiman 2001) for its ability to select important features, resist overfitting the data, model nonlinear relationships, handle categorical variables, and produce probabilistic output. These strengths, along with a record of attaining very high classification accuracy relative to other algorithms have led to widespread use of Random Forest in the astronomy community (e.g., Bailey et al. 2007; Carliles et al. 2010; Dubath et al. 2011; O'Keefe et al. 2009; Richards et al. 2011). In this work, we utilized custom `R` software built around the `randomForest` package to generate classifiers and evaluate performance.

Random Forest is an ensemble classifier that averages together the outputs from many decision trees, a common example of which is Classification and Regression Trees (CART, Breiman 1984). In RF, the decision trees are constructed by recursive binary splitting of the high-dimensional feature space, where each split is performed with respect to a particular feature. For example, the decision tree might split the data on feature $S/N_{\max}$ using value

---

[10]This of course would not be an issue when applying the `RATE` methodology to a problem with more well-defined class boundaries, such as prioritizing follow-up of a particular rare class of transient event.

[11]Bursts with a UVOT detection must be $z < 5$ due to the Lyman cutoff. This is due to the fact that photons with wavelengths smaller (thus higher energy) than the Lyman limit of $\lambda = 912$Å would be almost completely absorbed by neutral gas in the host galaxy and intergalactic star forming regions. A redshift of $z = 5$ might therefore be considered a natural cutoff point for the high-z class, but due to so few training events at this high redshift ($N_{z>5} = 8$), we opted for the more conservative cutoff point of $z = 4$ ($N_{z>4} = 18$).

100, in which case all observations with $S/N_{\max} > 100$ are placed in one group and the rest placed in the second group. As these are binary splits, for convenience we henceforth refer to observations going "left" or "right" of each split as an analogue for the decision made at that split.

For each split, the feature and specific split-point are chosen so as to best separate the observations into the classes, by using some objective function. We use the Gini Index, a standard objective function for classification (Breiman 1984). At any given node in a tree and some proposed split s, let $N_{l,h}$ = number of high-priority (in our case, high-z) events that go to the left of the split, $N_{l,l}$ = number of low-priority events that go left. Define $N_{r,h}$ and $N_{r,l}$ similarly, replacing left with right. Let $N_l = N_{l,l} + N_{l,h}$, the total number of observations that go left. Similarly define, $N_r = N_{r,l} + N_{r,h}$, for the total number of observations that go right. The Gini criterion is defined as

$$\frac{N_l}{N_l + N_r}\left(\frac{N_{l,h}}{N_l}\right)\left(\frac{N_{l,r}}{N_l}\right) + \frac{N_r}{N_l + N_r}\left(\frac{N_{r,h}}{N_r}\right)\left(\frac{N_{r,l}}{N_r}\right), \tag{1}$$

and the split that minimizes this value over the random subset of features considered at each node[12] is chosen. For instance, in the ideal case where the split on a particular feature completely separates all the instances of the two classes from each other, the Gini index reaches a minimum of 0. The splitting is done recursively, continuing down each subgroup until all of the observations in each final group ("terminal node") are of a single class. The process is known as "growing a tree" because each split can be visualized as generating two branches from a single branch to produce a tree-like structure. Once a tree is constructed from the training data, each new observation starts at the root node (the top split in the tree) and, recursively, the splitting rules determine the terminal node to which the observation belongs. The observation is assigned to the class of the terminal node.

To create the RF classifier, a sufficiently large[13] number of decision trees are constructed, resulting in a "forest". Each decision tree is generated from an independent bootstrap sample (Efron 1982); Samples are drawn with replacement from the original data set, resulting in a new data set of the same size as the original, with on average 2/3 of the original observations present at least once. Additionally, only a random subset of the features is eligible for splitting at each node. Many decision trees are grown with each tree slightly different due

---

[12]At each node, $m = 3$ features were considered, guided by the default practice in the `randomForest` routine of $m = \text{floor}(\sqrt{p})$, where p is the total number of features.

[13]With enough trees, error rates will converge and growing additional trees will result in no further performance improvements. Our forests are grown to 5000 trees throughout this work in order to ensure consistency in the rankings of unknown events.

to the bootstrap sampling and random selection of features at each split. RF classifies new observations by averaging the outputs of each tree in the ensemble.

Training observations can be classified by using all trees where that observation was not used in the bootstrap sampling stage. This produces estimates of error rates and class probabilities for each observation that are not overfit to the training data. Error rates and probabilities computed using this method are known as "out-of-bag" estimates.

3.1.1. Missing feature values

As mentioned in §2, certain features, namely α and $N_{\rm H,pc}$, were occasionally unable to be determined from model fits to the data and are thus missing for certain observations. We handle missing values by imputation, where missing values for features are assigned estimated values. For missing values of continuous features, we assigned the median of all observations for which that feature is non-missing. Missing categorical features are assigned the mode of all observations for which the feature is non-missing. This is one of the simplest imputation methods and has the advantage of being transparent and computationally cheap. We experimented with a more sophisticated imputation method, `MissForest`, that iteratively predicts the missing values of each feature given all the other features (Stekhoven & Bühlmann 2011), but as it produced similar error rates to median imputation, we opted for latter, simpler approach in our final classifier.

3.1.2. Class imbalance

A further challenge in this data set is the imbalance between classes. We are training on 135 bursts, only 18 of which are in the high-z class — an asymmetry present in many resource allocation problems where the goal is to prioritize the rarer events. Without modification, standard machine learning classification algorithms applied to imbalanced data sets attain notoriously suboptimal performance (Chawla et al. 2004), and often result in simply classifying all unknown events as the more common class. As we care more about correctly classifying the rarer events, misclassifications of high-z events must be punished more strongly than vice versa. In Random Forest, classes may be weighted in order to overcome the imbalance by altering the splits chosen by Gini and the probabilities assigned to classes in the terminal nodes of each tree (Chen et al. 2004).

We utilized the `classwt` option in the `randomForest` package, which accounts for class weights in the Gini index calculation (Eq. 1) when splitting at the nodes (Liaw 2011, private

communication), similar to weighting techniques used in single CART trees (Breiman 1984). If we are weighting high-priority observations (e.g. $z > 4$ GRBs) by w_h and low-priority observations by w_l, we let,

$$N'_{l,h} = w_h N_{l,h}$$
$$N'_{l,l} = w_l N_{l,l}$$
$$N'_{r,h} = w_h N_{r,h}$$
$$N'_{r,l} = w_l N_{r,l}$$

Let $N'_l = N'_{l,l} + N'_{l,h}$, the weighted total number of observations that go left. Similarly define, $N'_r = N'_{r,l} + N'_{r,h}$, for the weighted total number of observations that go right. The Gini criterion (Eq. 1) is evaluated with the weighted values, and the split that minimizes this value is chosen. We tested a variety of weight choices by fixing w_l to be unity and varying w_h over a range of values. The results of this test are presented in §4.1, which demonstrates the effects of class weight choice on classifier performance.

3.2.   RATE GRB-z : Random forest Automated Triage Estimator for GRB redshifts

With the background above in hand, we now describe our resource allocation algorithm and its utility for the prioritization of high-z GRB follow-up. In our application, the data are described in §2 and the classes are high- and low-redshift GRBs, with $z = 4$ as the boundary between the classes. Our primary goal is to provide a decision for each new GRB: should we devote further resources to this event or not? This decision may be different for each astronomer, as it is dependent on the amount of follow-up time available. Implicit in this goal is the desire to follow up on as many truly high-redshift bursts as possible, under a set of given telescope time constraints. Directly using the results of an off-the-shelf classifier for this task (i.e., strictly following-up on events labeled as "high-priority") is suboptimal. If too few events are labeled as high-priority, there would be an under-utilization of available resources. If too many are being labeled as high priority, simply following up on the first ones available would preclude any prioritization of events within this high-priority class.

These issues can be avoided by instead tailoring the follow-up decision to the resources available (in this case, the available telescope time devoted to high-z GRB observations). The RATE method works as follows: Let $\mathcal{Q}$ be the fraction of events one has resources to follow up on[14]. First we construct a Random Forest classifier using the training data with

---

[14]As telescope resources are allocated by number of hours and not number of objects, we implicitly assume

known response (in this case redshift). We compute the probability of each training event being high-priority using out-of-bag probabilities (See §3.1). For each new event, we obtain a probability of it being high priority using the Random Forest classifier, and compute the fraction of training bursts that received a higher probability of being high-priority than this new burst. A new burst is assigned rank n, with $n - 1$ training events having a lower probability of being high priority. Then, for N total training bursts, we obtain a learned probability rank for the new event of $\widehat{\mathcal{Q}} := n/(N+1)$. This leads to a simple decision metric for each new event: If $\widehat{\mathcal{Q}}$ is less than the desired fraction of events a particular observer wishes to follow up ($\widehat{\mathcal{Q}} < \mathcal{Q}$), follow-up observations are recommended. For instance, if one can afford to follow up on $\sim 30\%$ of all observable GRBs, then the desired follow-up fraction is $\mathcal{Q} = 0.3$, and follow-up would be recommended for all events assigned a $\widehat{\mathcal{Q}} < 0.3$. An illustration of this process in action is shown in Figure 3. The desired fraction of follow-up events $\mathcal{Q}$ can be dynamically changed without penalty; if the amount of available resources changes, one simply needs to raise or lower this cut-off value accordingly.

4.   Validation of Classifier Performance

Our training data consist of 135 bursts, 18 of which are high-redshift (> 4). Our primary measure of performance is efficiency, defined here as the fraction of high bursts that we that we follow up on relative to the number of total high-z GRBs that occurred ($N_{\text{high observed}}/N_{\text{total high}}$). A secondary performance measure is purity, the number of followed-up events that were actually high-z ($N_{\text{high observed}}/N_{\text{total observed}}$). We measure performance using 10-fold cross-validation (Kohavi 1995), where 90% of the data is used to construct a classifier and predict on the remaining 10% of events. Each line in the following performance plots is the cross-validated performance averaged across 100 trials of 10-fold cross-validation in order to reduce variability due to randomness in training/test subset selection.

4.1.   Comparison of Weight Choices

As described in §3.1.2, one of the primary challenges in learning on this dataset is the simple fact that there are comparatively few high-z events on which to train. If simply getting the most classifications correct were the primary performance metric, as it is in

---

here that an equal amount of resource time will be allocated to each follow-up event. This is not in general the case, as objects that turn out to be particularly interesting may have additional resources spent on them. However, a user's estimate of $\mathcal{Q}$ can always be adjusted without penalty as available resources change.
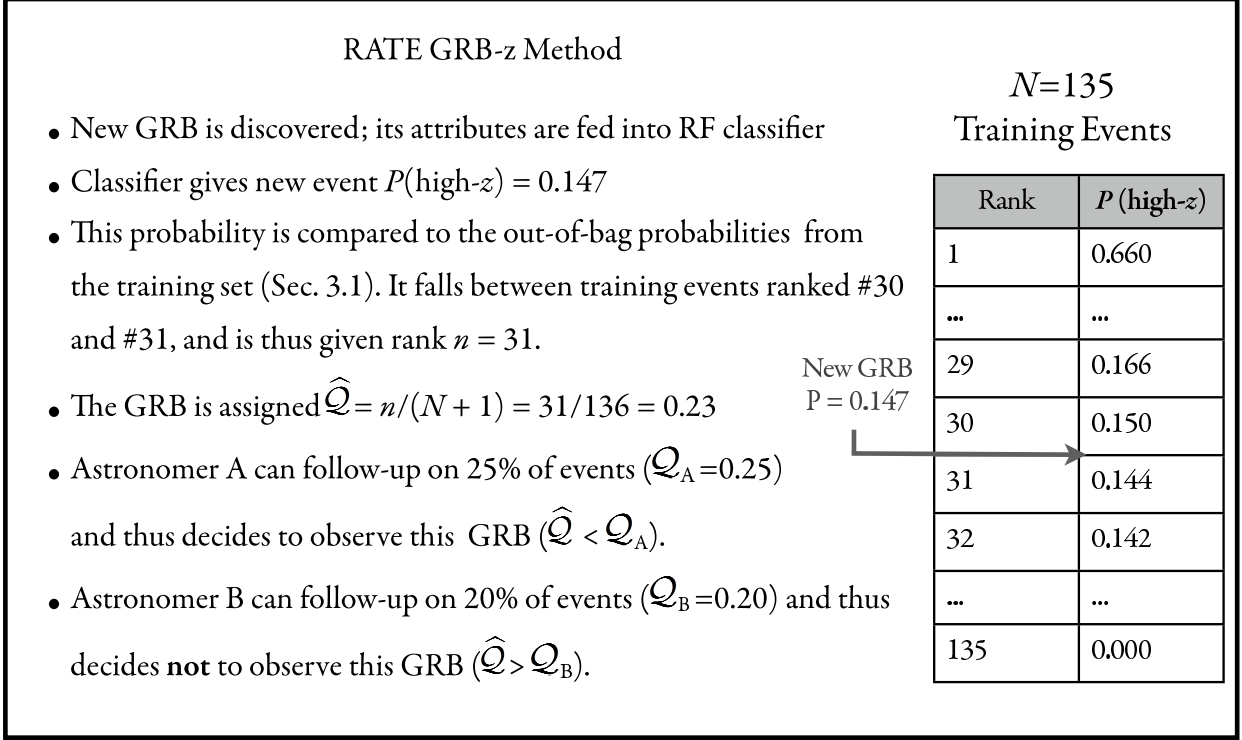
RATE GRB-z Method

N=135
Training Events

- New GRB is discovered; its attributes are fed into RF classifier
- Classifier gives new event $P(\text{high-}z) = 0.147$
- This probability is compared to the out-of-bag probabilities from the training set (Sec. 3.1). It falls between training events ranked #30 and #31, and is thus given rank $n = 31$.
- The GRB is assigned $\widehat{\mathcal{Q}} = n/(N + 1) = 31/136 = 0.23$
- Astronomer A can follow-up on 25% of events ($\mathcal{Q}_A = 0.25$) and thus decides to observe this GRB ($\widehat{\mathcal{Q}} < \mathcal{Q}_A$).
- Astronomer B can follow-up on 20% of events ($\mathcal{Q}_B = 0.20$) and thus decides **not** to observe this GRB ($\widehat{\mathcal{Q}} > \mathcal{Q}_B$).

New GRB
$P = 0.147$

Rank	P (high-z)
1	0.660
...	...
29	0.166
30	0.150
31	0.144
32	0.142
...	...
135	0.000

Fig. 3.— Example of the RATE GRB-z process.

many classification problems, classifying *all* new events as low-redshift would be considered a strong classifier since so few events are in the high-z class. However, since our objective is to identify the best candidates of this rare class, we punish misclassifications of high-z GRBs more heavily to achieve higher efficiency and purity (outlined above) for a given fraction of followed-up events.

Thus, in selecting the best weight for our classifier, we compared the efficiency and purity of high-z classification for various choices of the weight w_h using the feature set shown in Table 1. While the relative probability ranking of the GRBs stayed relatively stable over weight choices (Figure 4), a clear trend emerges when comparing classification performance (Figure 5). As expected, punishing misclassifications of the smaller, more desirable high-z class cause more of these rare events to be correctly identified. Beyond a weight of 10, however, a ceiling is reached where further weight increases show zero change in classification performance. This is therefore the weight chosen for all subsequent performance comparisons.
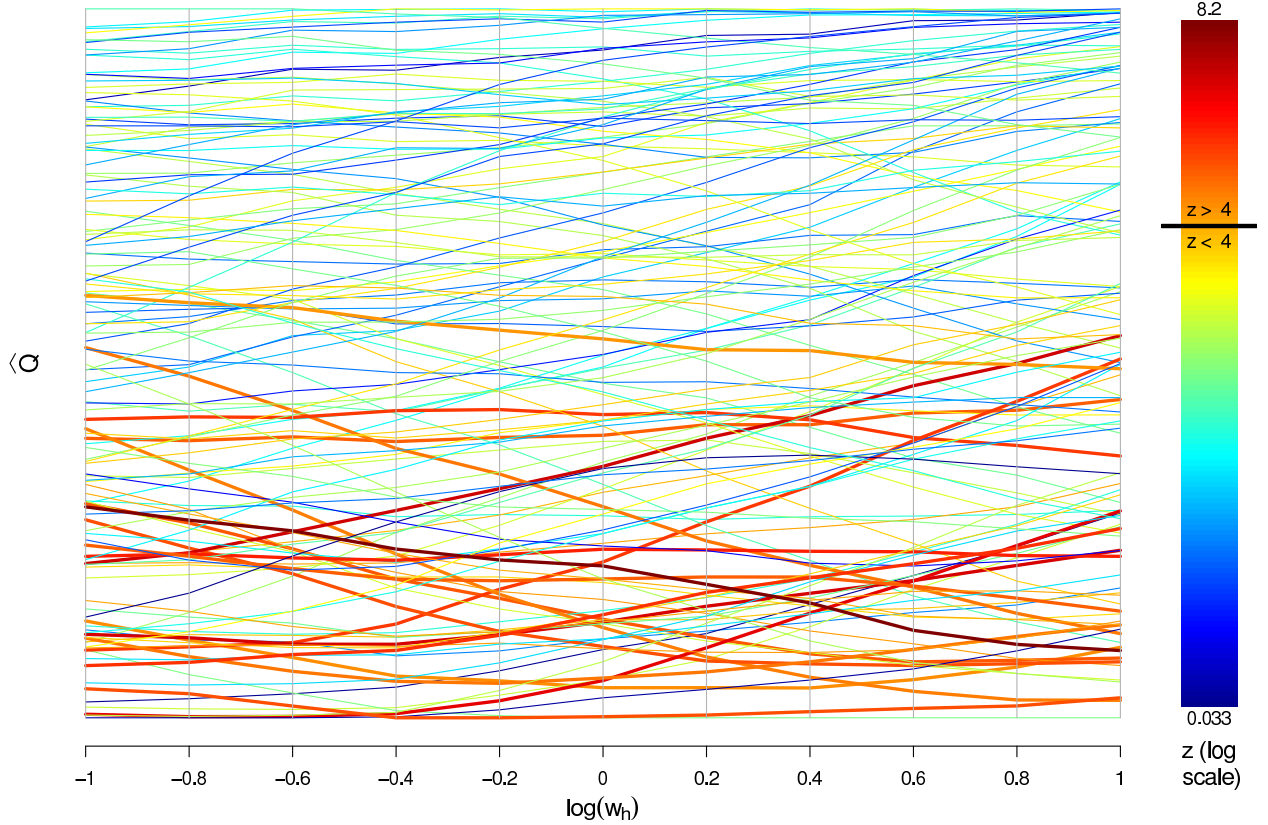
Fig. 4.— Bumps plot showing the cross-validated ranking prediction $\widehat{\mathcal{Q}}$ for each GRB in the training set over a variety of weight choices. Each line corresponds to an individual GRB, colored by its observed redshift. Bursts with $z > 4$ are plotted with a thicker line. The clustering of high-z events towards low $\widehat{\mathcal{Q}}$ is clear, illustrating the predictive power of the classifier. The relative ranking of events remains largely stable over different penalization weights, but performance improvements at higher weights are apparent in Figure 5, which level off after a weight of 10.

4.2. Effects of Feature Selection

As mentioned in §2, early testing indicated that the addition of too many features rapidly degraded the predictive power of the final classifier. This is due to a manifestation of the so-called "curse of dimensionality" known as Hughes Phenomenon (Hughes 1968), where for a fixed number of training instances, the predictive power decreases as the dimensionality increases. This appears to contradict the conventional wisdom that Random Forest does not overfit, and thus it is better to use many features. However, we note that resistance to overfitting is different from signal being drowned in noise. With enough noisy features,
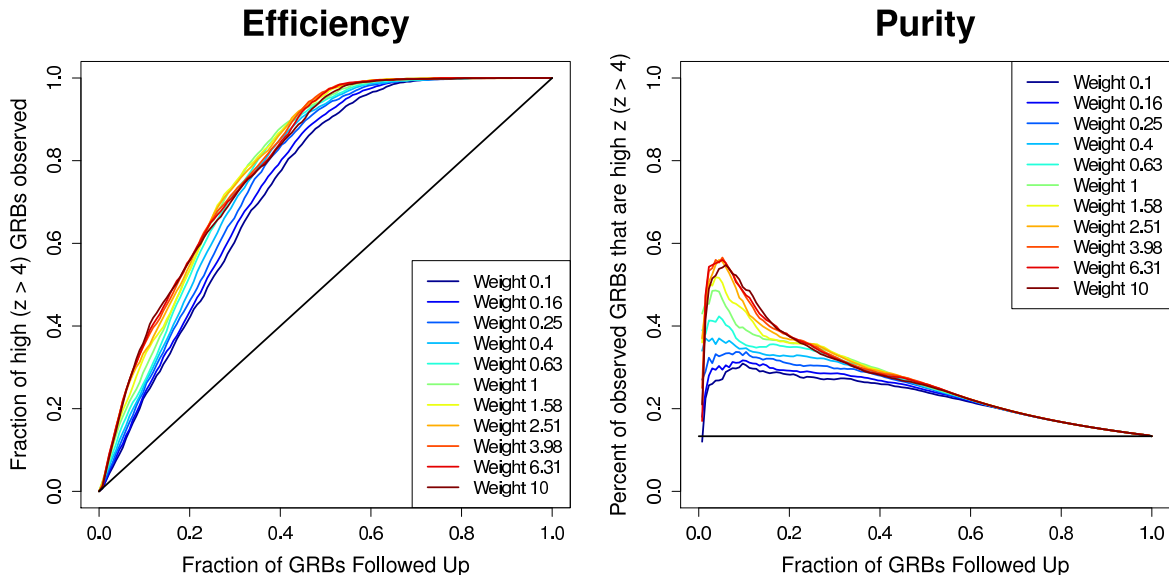
Fig. 5.— The effects of different weights on classifier performance are shown via plots of efficiency ($N_{\text{high observed}}/N_{\text{total high}}$; left panel) and purity ($N_{\text{high observed}}/N_{\text{total observed}}$; right panel) versus fraction of GRBs followed-up (see Figure 11 for how our decision criterion $\mathcal{Q}$ corresponds to actual fraction followed-up). Solid black lines show expected results if selecting events by random guessing alone. Cross validated performances of the classifier trained with different weights are shown. Weights above 1.0 penalize misclassifications of high-z events more strongly, and vice versa. Efficiency and purity were calculated at each fraction of followed-up GRBs ($\mathcal{Q}$, broken down into $N = 135$ bins) and averaged over 100 random number generator seeds to account for variance between Random Forest runs. Clear performance increases for both metrics are shown for higher weights, but beyond a weight of 10, identical results are achieved. For clarity, estimates of uncertainties in the curves are not shown, but are of order those plotted in Figure 7.

correlations between class and a useless feature will happen purely by chance, preventing true relationships from being found.

To visualize this effect for our data, we took our nominal feature set and continually added features with no predictive power (random samples from the uniform distribution) to quantify the degradation in performance of the resultant classifiers. The random features were re-generated for each of the 100 trials, and the cross-validated results are shown in Figure 6. The fact that even a small number of useless features causes a noticeable decrease in performance highlights the importance of attribute selection. However, we note that too much fine tuning of attribute feature selection choices — such as testing all combinations of
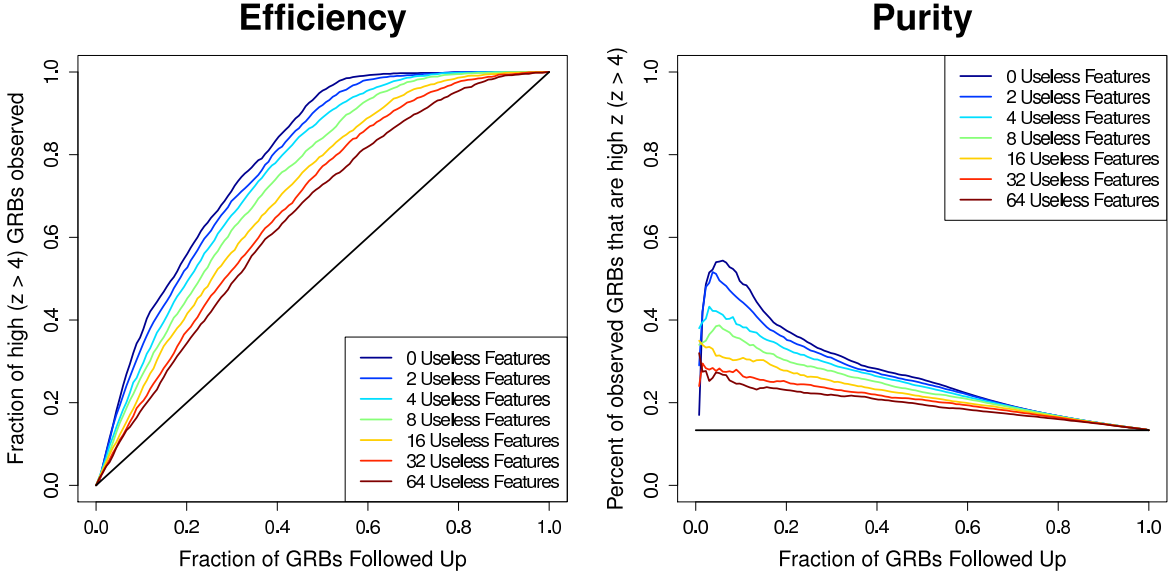
Fig. 6.— The effects of the addition of useless features on classifier performance are shown via plots of efficiency ($N_{\text{high observed}}/N_{\text{total high}}$; left panel) and purity ($N_{\text{high observed}}/N_{\text{total observed}}$; right panel) versus fraction of GRBs followed up according to our decision criterion ($\mathcal{Q}$). Solid black lines show expected results if selecting events by random guessing alone. Cross validated performances of the classifier trained with different amounts of useless, randomly generated features are shown. Degradation in both efficiency and purity becomes clear with the addition of only a few useless features, highlighting the importance of feature selection for small, imbalanced datasets such as this one.

features and seeing which one gives the best performance — would overfit to the data and give an underestimate of the true error.

4.3. Final Classifier

Taking into account the above issues of multiple feature set choices, the deleterious effect of useless features, and the performance with various weight choices to help with imbalance, we have developed a classifier which we believe to be robust and powerful. The full feature set utilized is shown in Table 1, and the weight chosen is described in §4.1. The final cross-validated estimates of $\widehat{\mathcal{Q}}$ for the training data are shown alongside the corresponding redshifts in Table 2. By referencing a particular point on the x-axis of Figure 7 (left panel) one can determine what fraction of high bursts can be detected for a particular amount of telescope

follow-up time. For example, if we are able to follow up on 20% of all GRBs detected by *Swift*, then the bursts recommended for follow-up by our classifier will contain on average $56\% \pm 6\%$ of all GRBs with redshift greater than 4 that occur. Following-up on $\sim 40\%$ of all bursts will yield $84\% \pm 6\%$ of all GRBs with redshift greater than 4, and following-up on the top 50% of candidates will result in nearly all of the high-z events being observed $(96\% \pm 4\%)$.

Purity is shown in the right panel of Figure 7, which describes how many of the followed-up bursts will actually be high-redshift. Following up on 20% of all bursts would result in $37\% \pm 4\%$ of the followed-up events being high-redshift, and $28\% \pm 2\%$ of followed up bursts would be high-redshift if 40% of GRBs were followed-up on.

As the high/low class division of $z = 4$ was relatively arbitrary, for completeness we also re-trained the classifier and calculated performance results using cutoff values of $z = 3.5$ (Fig. 8) and $z = 3$ (Fig. 9). Note that while the sample size of 'high' events more than doubles by lowering the cutoff value to $z = 3$, the resultant efficiency decreases significantly. We attribute this effect to a decrease in the predictive power of certain attributes at lower redshift. For instance, the $z > 3$ population has proportionally many more instances of UVOT detections in its 'high-z' class than the $z > 4$ population, which reduces its effectiveness as a discriminating feature.

4.4. Feature Importance

There are several complications in identifying the relative importance of features in contributing to selecting high-z candidates. To an extent, simple scatter plots such as those in Figure 2 can give an indication as to what features are best at separating the classes, but these fail to account for the complex interactions between features occurring within the RF classification. The effects of removing features from the dataset and then re-constructing the classifier give another indication of feature importance, but fail to account for redundancy in the features; if two features have similar predictive properties, removing one will just cause the other to take its place. Nevertheless, such an experiment can be illustrative, and the results are shown in Figure 10. In general, the removal of an individual feature does not cause a significant change in performance, and the small changes that do occur trend toward a degradation in the number of high-z bursts identified, implying that few if any of the features in the dataset are useless. The features that cause the largest degradation in performance upon their removal are α, $R_{peak,BAT}$, and $S/N_{\max}$, indicating that these features are both useful predictors and are not fully redundant with other features. Note that the slight improvement in performance from the removal of the temporal features T_{90} and t_{BAT}
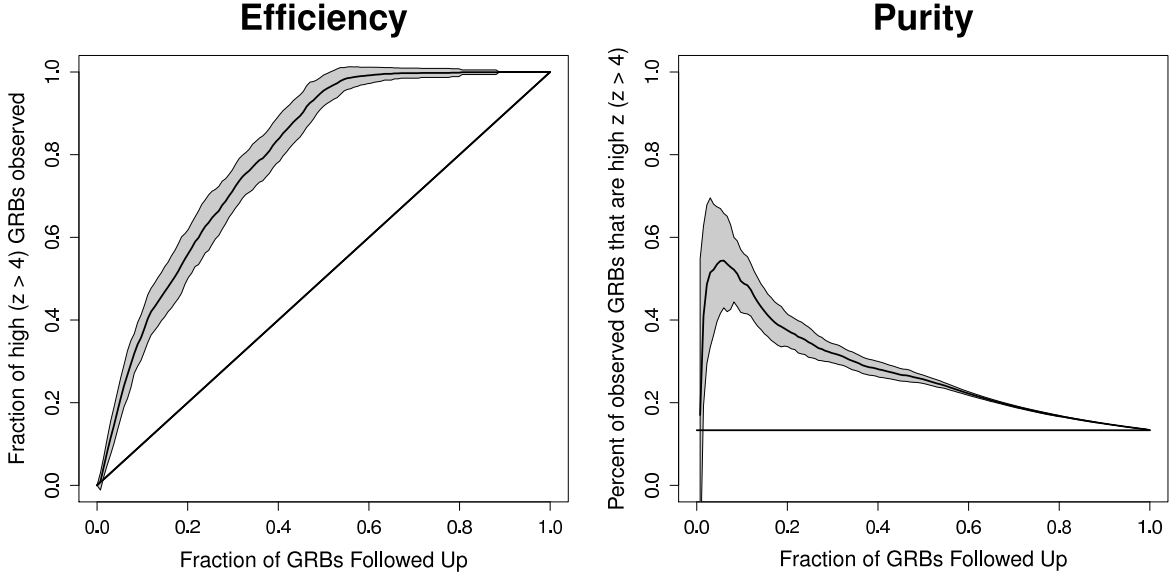
Efficiency

Purity

Fig. 7.— Efficiency $(N_{\text{high observed}}/N_{\text{total high}};$ left panel) and purity $(N_{\text{high observed}}/N_{\text{total observed}};$ right panel) versus fraction of GRBs followed up according to our decision criterion $(\mathcal{Q})$ with a high-z cutoff of $z = 4$. 18 bursts ($\sim 13\%$ of our training set) are $z \geq 4.0$. The curve uncertainties shown are 1σ standard deviations from the mean value across all seeds.

is consistent with these values having little-to-no predictive power, in agreement with the recent findings of Kocevski & Petrosian (2011) showing a lack of time dilation signatures in GRB light curves.

5. Discussion

5.1. Calibration on GRBs with unknown redshifts

A natural application of our methodology is to use it to predict the follow-up metric $\widehat{\mathcal{Q}}$ for the remaining majority of long-duration *Swift* GRBs with no known redshift, providing a list of the top candidates predicted to be high-z. This application is precisely how RATE GRB-z could be used in practice on new events, albeit one-at-a-time rather than on many at once. We caution that due to the natural selection effect of GRBs with measured redshifts having a higher likelihood of being brighter events, the bursts with unknown redshifts are likely to comprise a somewhat different redshift distribution than our training dataset. The primary consequence of this is the interpretation of the user-desired follow-up fraction $\mathcal{Q}$
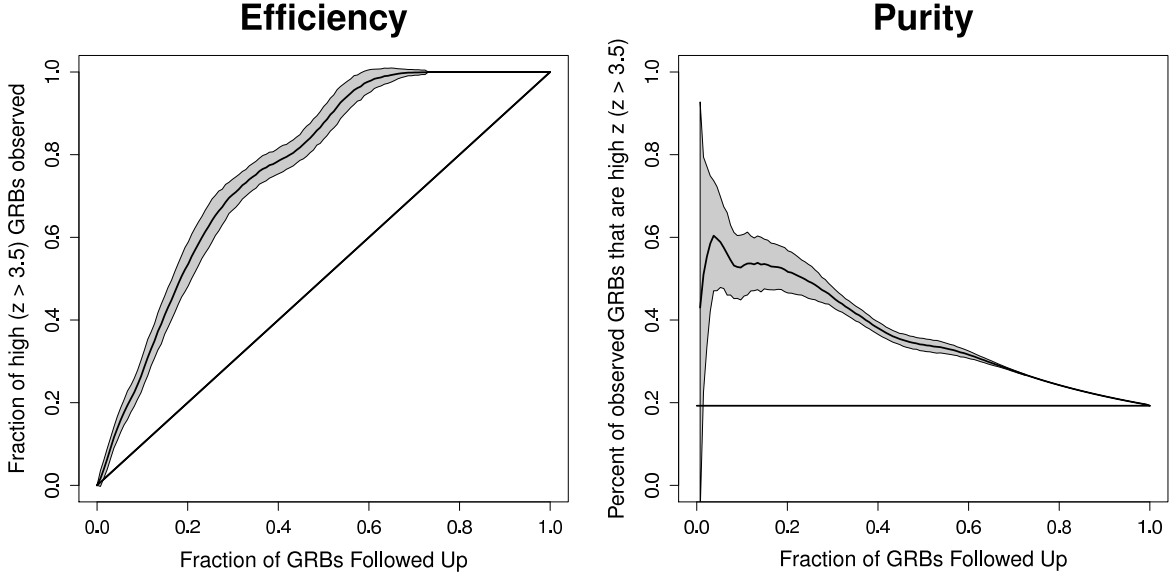
Fig. 8.— Efficiency ($N_{\text{high observed}}/N_{\text{total high}}$; left panel) and purity ($N_{\text{high observed}}/N_{\text{total observed}}$; right panel) versus fraction of GRBs followed up according to our decision criterion ($\mathcal{Q}$)with a high-z cutoff of $z = 3.5$. 26 bursts ($\sim 19\%$ of our training set) are $z \geq 3.5$. The curve uncertainties shown are 1σ standard deviations from the mean value across all seeds.

and the prioritization parameter $\widehat{\mathcal{Q}}$. In principle, the classifier was calibrated such that, over time, a fraction $\mathcal{Q}$ of new events will have affirmative follow-up recommendations (that is, events such that $\widehat{\mathcal{Q}} \leq \mathcal{Q}$). However, this will not necessarily be the case if the full redshift distribution of GRBs makes up a different population than our training data.

To test this, we calculated $\widehat{\mathcal{Q}}$ for each of the remaining 212 GRBs with unknown redshift that met our culling criteria outlined in §2. From this we could calculate the fraction of GRBs followed up ($\widehat{\mathcal{Q}} \leq \mathcal{Q}$) for each cutoff value of $\mathcal{Q}$. The results of this test are shown in Figure 11. For the chosen weight of 10 (see §4.1), the $\mathcal{Q}$-values are well calibrated with the final follow-up recommendations. The resultant $\widehat{\mathcal{Q}}$ priorities are listed in Table 4. These values can be interpreted as a ranking of which of these past events without secure redshift determinations are most likely to be at high-redshift.
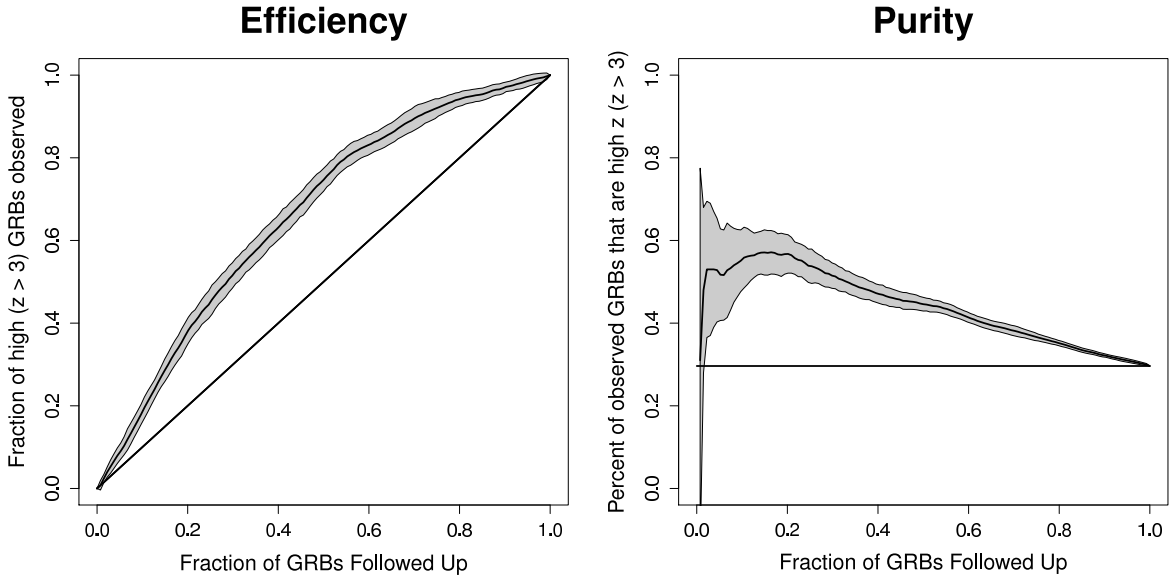
Fig. 9.— Efficiency ($N_{\text{high observed}}/N_{\text{total high}}$; left panel) and purity ($N_{\text{high observed}}/N_{\text{total observed}}$; right panel) versus fraction of GRBs followed up according to our decision criterion ($\mathcal{Q}$)with a high-z cutoff of $z = 3.0$. 40 bursts ($\sim 30\%$ of our training set) are $z \geq 3.0$. The curve uncertainties shown are 1σ standard deviations from the mean value across all seeds.
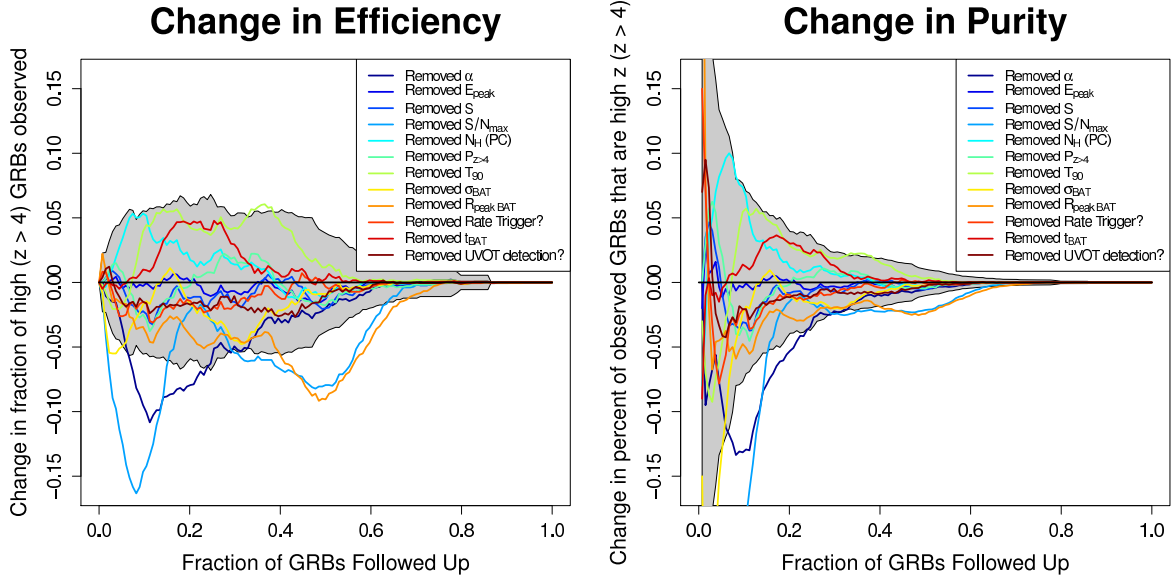
Fig. 10.— Change in efficiency (left) and purity (right) by removing individual features from the default feature set listed in Table 1. The standard deviation from the mean value across all seeds for the default dataset is shown in grey. The lack of degradation in performance by the removal of a feature does not necessarily imply that it has no predictive power, only that it may be redundant with other features. Most of the features do not cause a significant change in performance once removed from the dataset. However, the removal of a few of the individual features does cause a degradation in performance larger than what would be expected by random, implying that these features are both important and not completely redundant. Note that the relative change in both purity and efficiency are equal in both plots, as only the numerator of each metric is changing ($N_{\mathrm{high\ observed}}$), but we show both values for consistency.

Table 4. Test Data

GRB	$\widehat{\mathcal{Q}}$	α	E_{peak} (keV)	S (erg/cm^2)	S/N_{max}	$N_{H,pc}$ (10^{22} cm^{-2})	T_{90} (s)	σ_{BAT}	$R_{peak,BAT}$ (ct/s)	Rate trigger	t_{BAT} (s)	UVOT detect	$P_{z>4}$
050215A	3.19e-01	-1.29e+00	4.14e+02	1.34e-06	1.02e+01	?	6.65e+01	9.00e+00	6.94e+02	yes	8.19e+00	no	9.81e-02
050215B	1.78e-01	?	3.01e+01	2.86e-07	1.44e+01	5.70e-02	8.50e+00	8.00e+00	3.00e+02	yes	2.05e+00	no	1.06e-01
050219A	4.22e-01	1.87e-02	1.00e+02	4.91e-06	5.08e+01	9.10e-02	2.50e+01	8.00e+00	1.93e+02	yes	1.02e+00	no	1.12e-01
050219B	7.33e-01	-8.94e-01	1.12e+02	1.94e-05	7.19e+01	8.80e-02	2.09e+01	1.70e+01	4.09e+02	yes	1.02e+00	no	2.73e-02
050326	7.04e-01	-1.04e+00	3.41e+02	1.70e-05	1.33e+02	3.80e-02	3.02e+01	2.10e+01	1.84e+04	yes	5.12e-01	no	5.67e-02

Note. — Table 4 is published in its entirety in the electronic edition of The Astrophysical Journal. A portion is shown here for guidance regarding its form and content.
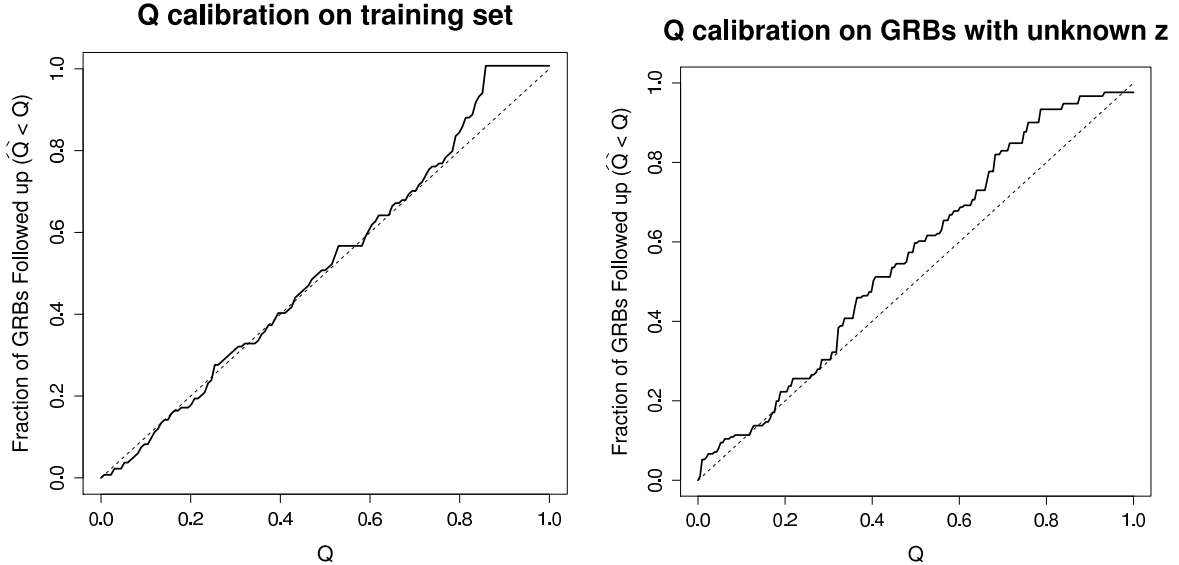
Fig. 11.— Here we quantify the calibration of $\mathcal{Q}$; namely, how well does the user-desired follow-up fraction $\mathcal{Q}$ correspond to the actual number of bursts recommended to be followed up by the algorithm ($\widehat{\mathcal{Q}} \leq \mathcal{Q}$). The left figure shows the self-calibration of the cross-validated training set, which aligns as expected. The right plot shows the calibration on the test set is good, especially at low $\mathcal{Q}$. At larger $\mathcal{Q}$, there is a slight departure from the diagonal, implying a follow-up recommendation of more events than expected at these values. This can be attributed to the differing populations between the training set (with measured redshifts) and test set (with unknown redshifts), as illustrated in Figure 2. This slight discrepancy is not surprising, as low brightness events without UVOT detections are naturally more difficult to obtain redshifts for.

5.2. Validation Set: Application to Recent GRBs

Since the cutoff date in our training set (June 21, 2010) until Sept. 1, 2011, there have been 15 long duration *Swift* GRBs with reliable redshifts from which we constructed an independent validation set to test our method[15]. The feature values for these GRBs are presented in Table 5. While none of these events were over our high-redshift cutoff value of $z = 4$, it is still possible, though challenging, to use low-z events (either by direct redshift measurement or by the identification of a coincident blue host galaxy) as a consistency

---

[15]One of the bursts with a measured redshift, GRB 110328A, had very unusual properties and was determined to be a potential Tidal Disruption Event (Bloom et al. 2011; Levan et al. 2011b), and was thus also excluded from the validation set.

test. We would expect that the purity at a given $\mathcal{Q}$ would be lower than the fraction of recommended follow-up events ($\widehat{\mathcal{Q}} \leq \mathcal{Q}$) *without* a secure low-z determination. For instance, $\mathcal{Q} = 0.2$ has a purity of $37\% \pm 4\%$, so no more than $\sim 63\%$ of events with $\widehat{\mathcal{Q}} < 0.2$ should be definitively low-redshift.

The validation GRBs were run through the `RATE GRB-`z classifier, and their resultant $\widehat{\mathcal{Q}}$ values are shown in Table 6 along with their corresponding redshifts. The smallest $\widehat{\mathcal{Q}}$ value of these events is ~ 0.3, meaning that none of these events would have been recommended for high-z follow-up for anyone wishing to observe fewer than 30% of events. While these values are certainly consistent with our expected purity, it is not particularly constraining, as it would have been very unlikely for this almost-random selection of GRBs to violate this constraint by chance alone, even if the classifier had no predictive power.

A more constraining test is the identification of high-z events with high $\widehat{\mathcal{Q}}$ for comparison with the expected efficiency. Two events not included in our training set have had recent high-z identifications: GRB 090429B with strong photometric evidence for being $z \simeq 9.4$ (Cucchiara et al. 2011b), and the spectroscopic identification of GRB 111008A at $z = 4.99$ (Levan et al. 2011a; Wiersema et al. 2011). The former has a $\widehat{\mathcal{Q}}$ value of ~ 0.185, consistent with the expected efficiency. However, GRB 111008A has a $\widehat{\mathcal{Q}}$ of ~ 0.637, a value above which we would have expected to find no more than 1% of high-z events. This outlier seems likely due to the extreme brightness of the event (among the brightest $\sim 10\%$ of *Swift* bursts in the observer frame, and top $\sim 3\%$ in the rest frame). Indeed, compared to all 18 high-z events in the training set, GRB 111008A has the most extreme values towards the 'wrong' end of three of the highly important features identified in §4.4 (α, $P_{z>4}$, and $R_{peak,BAT}$) and also has the fourth largest S/N_{max}. In later iterations of `RATE GRB-`z, this event (and all new GRBs with secure redshifts) will be added to the training data to re-generate the classifier and further improve its robustness against such outliers.

Table 5. Validation Data

GRB	$\widehat{\mathcal{Q}}$	α	E_{peak} (keV)	S (erg/cm^2)	S/N_{max}	$N_{H,pc}$ (10^{22} cm^{-2})	T_{90} (s)	σ_{BAT}	$R_{peak,BAT}$ (ct/s)	Rate trigger	t_{BAT} (s)	UVOT detect	$P_{z>4}$
100728B	6.07e-01	-1.64e+00	8.19e+01	2.54e-06	2.06e+01	3.90e-02	1.15e+01	9.07e+00	1.47e+02	yes	1.02e+00	yes	1.01e-01
100814A	6.81e-01	-1.11e+00	1.35e+02	9.33e-06	9.80e+01	?	1.77e+02	1.91e+01	8.34e+02	yes	1.02e+00	yes	1.80e-01
100816A	9.33e-01	-5.71e-01	1.42e+02	2.71e-06	5.80e+01	1.13e-01	2.50e+00	2.29e+01	1.42e+03	yes	1.02e+00	yes	5.55e-02
100901A	4.00e-01	-1.55e+00	1.28e+02	3.41e-06	1.78e+01	4.00e-02	4.59e+02	7.70e+00	4.50e+02	yes	8.19e+00	yes	2.25e-01
100906A	1.00e+00	-1.66e+00	1.57e+02	1.37e-05	1.36e+02	?	1.17e+02	1.05e+01	1.91e+02	yes	5.12e-01	yes	7.39e-02
101219B	6.30e-01	-1.89e+00	4.97e+01	3.75e-06	1.00e+01	-8.00e-03	4.18e+01	7.63e+00	8.44e+02	no	6.40e+01	yes	1.07e-01
110205A	3.19e-01	-1.39e+00	9.75e+01	1.98e-05	1.50e+02	1.10e-02	2.77e+02	1.00e+01	1.48e+03	no	6.40e+01	yes	1.45e-01
110213A	9.33e-01	-1.82e+00	6.70e+01	8.77e-06	3.10e+01	4.00e-02	4.31e+01	1.21e+01	2.05e+02	yes	1.02e+00	yes	5.32e-02
110422A	1.00e+00	-6.23e-01	1.11e+02	5.17e-05	2.10e+02	1.58e-01	2.67e+01	7.19e+00	8.20e+01	yes	1.28e-01	yes	2.49e-02
110503A	9.33e-01	-8.18e-01	1.42e+02	1.43e-05	6.27e+01	2.60e-02	9.31e+00	2.04e+01	1.26e+03	yes	1.02e+00	yes	1.89e-02
110715A	9.33e-01	-1.06e+00	8.94e+01	1.40e-05	2.02e+02	1.64e-01	1.31e+01	1.19e+01	1.47e+02	yes	1.28e-01	yes	9.70e-03
110726A	5.04e-01	-2.97e-01	4.27e+01	2.07e-07	1.51e+01	-4.90e-02	5.40e+00	8.60e+00	2.24e+02	yes	1.02e+00	yes	1.14e-01
110731A	1.00e+00	-1.19e+00	4.06e+02	1.25e-05	1.30e+02	7.20e-02	4.66e+01	2.46e+01	2.32e+03	yes	1.02e+00	yes	5.09e-02
110801A	9.33e-01	-1.84e+00	6.07e+01	6.85e-06	3.56e+01	2.90e-02	4.00e+02	7.83e+00	3.50e+02	yes	4.10e+00	yes	1.98e-01
110808A	5.56e-01	?	2.59e+01	4.27e-07	1.01e+01	2.17e-01	3.94e+01	7.19e+00	4.26e+02	yes	8.19e+00	yes	1.06e-01

5.3. Comparison to Previous Efforts

Extracting indications of redshift from promptly available information has been a continuing goal of GRB studies since their cosmological origins were discovered nearly 15 years ago. Several potential luminosity indicators were pursued with the optimistic goal of using GRBs as standard candles for cosmological studies. The efficacy of individual indicators toward this goal proved to be limited, and a physical origin of the relations has been contested, with authors attributing them instead to detector thresholding or other selection effects (Butler et al. 2007, 2009, 2010; Shahmoradi & Nemiroff 2011). While these studies have ruled out the majority of such relations as intrinsic to GRBs themselves, prompt properties can still be used as redshift indicators if the systematics are properly accounted for.

Several recent studies have attempted to use combinations of features to determine "pseudo-redshifts" for GRBs. In an extension of work by Schaefer (2007), Xiao & Schaefer (2009, 2011) used a combination of six purported luminosity relations. Further, Koen (2009, 2010) has explored linear regression as a tool for predicting GRB redshifts using the dataset from Schaefer (2007). As data derived from multiple satellites were used, these studies are particularly vulnerable to the detector selection effects mentioned above.

Some works avoided the complications of regression and instead focused upon the simple selection of high-z candidates for follow-up purposes. Campana et al. (2007) utilized a sample of *Swift*-only bursts (thus avoiding detector effect biases) and used hard cuts on three features (T_{90}, lack of UVOT detection, and high-galactic latitude) for high-z candidate selection. Salvaterra et al. (2007) extended upon this work with the additional feature of peak photon flux.

Several issues prevent a direct comparison among the various methods of the effectiveness at separating high-z events. These include the usage of different features from each study, which is complicated by the lack of uniformity of features being created for each. Further, the techniques above strictly constrain the manner in which each feature influences the output, whereas our method is fully non-parametric and therefore more flexible. However, the largest concern is accurate reporting of predictive performance. In particular, we caution against the circular practice of measuring the performance of methods by applying them to the same events from which the luminosity relations were formed. In order to prevent overestimating the accuracy of a predictive model, one needs to test on data independent from the training set, such as with cross-validation.

Finally, the `RATE` method differs from previous efforts in that it casts the problem as one of optimal resource allocation under limited follow-up time. Prior techniques are not

explicitly calibrated to suit this purpose. Direct classification methods will either under or over-utilize available resources. Past regression or "pseudo-z" methods are not explicitly calibrated to a particular follow-up decision (i.e., at what "pseudo-z" does one decide to follow up?), though it would be possible in principle to correct for this using a transformation which ensures that the desired follow-up fraction corresponds to the actual fraction of bursts followed up (e.g., Figure 11). In contrast, the RATE technique is by design applicable to any available resource reserves, and is generally extendable to any transient follow-up prioritization problem.

6. Conclusions

In this paper, we presented the RATE GRB-z method for allocating follow-up telescope resources to high-redshift GRB candidates using Random Forest classification on early-time *Swift* metrics. The RATE method is generalizable to any prioritization problem that can be parameterized as "observe" or "don't observe", and accommodates statistical challenges such as small datasets, imbalanced classes, and missing feature values. The issue of resource allocation is becoming increasingly important in the era of data-driven transient surveys such as PTF, Pan-STARRS, and LSST which provide extremely high discovery rates without a significant increase in follow-up resources. With enough training instances of any object of interest for a given transient survey, the RATE method can be applied to prioritize follow-up of future high-priority candidates.

In the RATE GRB-z application, our robust, cross-validated performance metrics indicate that by observing just 20% of bursts, one can capture 56%±6% of $z > 4$ events with a sample purity of 37%±4%. Further, following up on half of all events will yield nearly all (96%±4%) of the high-z events. The method provides a simple decision point for each new event: if the prioritization value $\widehat{\mathcal{Q}}$ is smaller than the percent of events a user wishes to allocate resources to, then follow-up is recommended. These rapid predictions, combined with the more traditional photometric dropout technique from simultaneous multi-filter NIR observatories (such as PAIRITEL, GROND, and the upcoming RATIR), offer a robust tool in more efficiently informing GRB follow-up decisions. To facilitate the dissemination of high-redshift GRB predictions to the community, we have set up a website (`http://rate.grbz.info`) with $\widehat{\mathcal{Q}}$ values for past bursts, and an RSS feed (`http://rate.grbz.info/rss.xml`) to provide real-time results from our classifier on new events.

Facilities: Swift (BAT, XRT, UVOT)

REFERENCES

Amati, L., et al. 2002, A&A, 390, 81

Antonelli, L. A., et al. 2010, GRB Coordinates Network, 10620, 1

Bailey, S., Aragon, C., Romano, R., Thomas, R., Weaver, B., & Wong, D. 2007, The Astrophysical Journal, 665, 1246

Band, D. L., Norris, J. P., & Bonnell, J. T. 2004, ApJ, 613, 484

Barkana, R., & Loeb, A. 2004, ApJ, 601, 64

Barthelmy, S. D., et al. 2005, Space Sci. Rev., 120, 143

Berger, E. 2006, GRB Coordinates Network, 5962, 1

Berger, E., Cenko, S. B., Steidel, C., Reddy, N., & Fox, D. B. 2005, GRB Coordinates Network, 3368, 1

Berger, E., Foley, R., Simcoe, R., & Irwin, J. 2008a, GRB Coordinates Network, 8434, 1

Berger, E., Fox, D. B., & Cucchiara, A. 2007, GRB Coordinates Network, 6470, 1

Berger, E., Fox, D. B., Cucchiara, A., & Cenko, S. B. 2008b, GRB Coordinates Network, 8335, 1

Berger, E., & Gladders, M. 2006, GRB Coordinates Network, 5170, 1

Berger, E., Kulkarni, S. R., Rau, A., & Fox, D. B. 2006, GRB Coordinates Network, 4815, 1

Berger, E., & Mulchaey, J. 2005, GRB Coordinates Network, 3122, 1

Berger, E., & Rauch, M. 2008, GRB Coordinates Network, 8542, 1

Berger, E., & Shin, M.-S. 2006, GRB Coordinates Network, 5283, 1

Bloom, J. S., Perley, D. A., & Chen, H. W. 2006, GRB Coordinates Network, 5826, 1

Bloom, J. S., et al. 2011, Science, 333, 203

Bouwens, R. J., et al. 2010, ApJ, 709, L133

—. 2011, Nature, 469, 504

Breiman, L. 1984, Classification and Regression Trees (Chapman & Hall/CRC)

—. 2001, Machine Learning, 45, 5

Burrows, D. N., et al. 2005, Space Sci. Rev., 120, 165

Butler, N. R., Bloom, J. S., & Poznanski, D. 2010, ApJ, 711, 495

Butler, N. R., & Kocevski, D. 2007, ApJ, 663, 407

Butler, N. R., Kocevski, D., & Bloom, J. S. 2009, ApJ, 694, 76

Butler, N. R., Kocevski, D., Bloom, J. S., & Curtis, J. L. 2007, ApJ, 671, 656

Cabrera Lavers, A., de Ugarte Postigo, A., Castro-Tirado, A. J., Gorosabel, J., Thoene, C. C., & Dominguez, R. 2011, GRB Coordinates Network, 12234, 1

Campana, S., Tagliaferri, G., Malesani, D., Stella, L., D'Avanzo, P., Chincarini, G., & Covino, S. 2007, A&A, 464, L25

Carliles, S., et al. 2010, The Astrophysical Journal, 712, 511

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. 2006, Measurement Error in Nonlinear Models: A Modern Perspective, second edition edn. (Chapman & Hall/CRC)

Cenko, S. B., Berger, E., Djorgovski, S. G., Mahabal, A. A., & Fox, D. B. 2006, GRB Coordinates Network, 5155, 1

Cenko, S. B., Bloom, J. S., Perley, D. A., Cobb, B. E., Morgan, A. N., Miller, A. A., Modjaz, M., & James, B. 2010a, GRB Coordinates Network, 10389, 1

Cenko, S. B., Gezari, S., Small, T., Fox, D. B., & Chornock, R. 2007, GRB Coordinates Network, 6322, 1

Cenko, S. B., Hora, J. L., & Bloom, J. S. 2011, GRB Coordinates Network, 11638, 1

Cenko, S. B., Kulkarni, S. R., Gal-Yam, A., & Berger, E. 2005, GRB Coordinates Network, 3542, 1

Cenko, S. B., Perley, D. A., Junkkarinen, V., Burbidge, M., Diego, U. S., & Miller, K. 2009, GRB Coordinates Network, 9518, 1

Cenko, S. B., Perley, D. A., Morgan, A. N., Klein, C. R., Bloom, J. S., Butler, N. R., & Cobb, B. E. 2010b, GRB Coordinates Network, 10752, 1

Chawla, N. V., Japkowicz, N., & Kotcz, A. 2004, SIGKDD Explor. Newsl., 6, 1

Chen, C., Liaw, A., & Breiman, L. 2004, Technical Report

Chen, H.-W., Helsby, J., Shectman, S., Thompson, I., & Crane, J. 2009, GRB Coordinates Network, 10038, 1

Chen, H.-W., Prochaska, J. X., Bloom, J. S., & Thompson, I. B. 2005, ApJ, 634, L25

Chornock, R., Berger, E., Fox, D., Levan, A. J., Tanvir, N. R., & Wiersema, K. 2010a, GRB Coordinates Network, 11164, 1

Chornock, R., Cenko, S. B., Griffith, C. V., Kislak, M. E., Kleiser, I. K. W., & Filippenko, A. V. 2009a, GRB Coordinates Network, 9151, 1

Chornock, R., Cucchiara, A., Fox, D., & Berger, E. 2010b, GRB Coordinates Network, 10466, 1

Chornock, R., Perley, D. A., Cenko, S. B., & Bloom, J. S. 2009b, GRB Coordinates Network, 9243, 1

Chornock, R., Perley, D. A., & Cobb, B. E. 2009c, GRB Coordinates Network, 10100, 1

Chornock, R., et al. 2010c, ArXiv e-prints 1004.2262

Cucchiara, A., Bloom, J. S., & Cenko, S. B. 2011a, GRB Coordinates Network, 12202, 1

Cucchiara, A., Fox, D., Levan, A., & Tanvir, N. 2009, GRB Coordinates Network, 10202, 1

Cucchiara, A., & Fox, D. B. 2008, GRB Coordinates Network, 7654, 1

—. 2010, GRB Coordinates Network, 10624, 1

Cucchiara, A., Fox, D. B., & Berger, E. 2006a, GRB Coordinates Network, 4729, 1

Cucchiara, A., Fox, D. B., & Cenko, S. B. 2007a, GRB Coordinates Network, 7124, 1

Cucchiara, A., Fox, D. B., Cenko, S. B., & Berger, E. 2008a, GRB Coordinates Network, 8346, 1

—. 2008b, GRB Coordinates Network, 8448, 1

—. 2008c, GRB Coordinates Network, 8713, 1

—. 2008d, GRB Coordinates Network, 8065, 1

Cucchiara, A., Fox, D. B., Cenko, S. B., & Price, P. A. 2007b, GRB Coordinates Network, 6083, 1

Cucchiara, A., Marshall, F. E., & Guidorzi, C. 2007c, GRB Coordinates Network, 6419, 1

Cucchiara, A., Price, P. A., Fox, D. B., Cenko, S. B., & Schmidt, B. P. 2006b, GRB Coordinates Network, 5052, 1

Cucchiara, A., et al. 2011b, ApJ, 736, 7

D'Avanzo, P., D'Elia, V., & Covino, S. 2008a, GRB Coordinates Network, 8350, 1

D'Avanzo, P., Fiore, F., Piranomonte, S., Covino, S., Tagliaferri, G., Chincarini, G., & Stella, L. 2007, GRB Coordinates Network, 7152, 1

D'Avanzo, P., et al. 2008b, GRB Coordinates Network, 7997, 1

—. 2010, A&A, 522, A20+

de Ugarte Postigo, A., Castro-Tirado, A. J., & Gorosabel, J. 2011a, GRB Coordinates Network, 11978, 1

de Ugarte Postigo, A., Castro-Tirado, A. J., Tello, J. C., Cabrera Lavers, A., & Reverte, D. 2011b, GRB Coordinates Network, 11993, 1

de Ugarte Postigo, A., Gorosabel, J., Fynbo, J. P. U., Wiersema, K., & Tanvir, N. 2009a, GRB Coordinates Network, 9771, 1

de Ugarte Postigo, A., Gorosabel, J., Malesani, D., Fynbo, J. P. U., & Levan, A. J. 2009b, GRB Coordinates Network, 9383, 1

de Ugarte Postigo, A., Jakobsson, P., Malesani, D., Fynbo, J. P. U., Simpson, E., & Barros, S. 2009c, GRB Coordinates Network, 8766, 1

de Ugarte Postigo, A., Thoene, C. C., Vergani, S. D., Milvang-Jensen, B., & Fynbo, J. 2010, GRB Coordinates Network, 10445, 1

de Ugarte Postigo, A., et al. 2011c, GRB Coordinates Network, 11579, 1

—. 2011d, GRB Coordinates Network, 12258, 1

D'Elia, V., Covino, S., & D'Avanzo, P. 2008a, GRB Coordinates Network, 8438, 1

D'Elia, V., Thoene, C. C., de Ugarte Postigo, A., D'Avanzo, P., Covino, S., Piranomonte, S., Salvaterra, R., & Chincarini, G. 2008b, GRB Coordinates Network, 8531, 1

D'Elia, V., et al. 2005, GRB Coordinates Network, 4044, 1

—. 2006, GRB Coordinates Network, 5637, 1

—. 2007, A&A, 467, 629

Della Valle, M., et al. 2006, Nature, 444, 1050

Dubath, P., et al. 2011, MNRAS, 414, 2602

Efron, B. 1982, The jackknife, the bootstrap, and other resampling plans. (Society of Industrial and Applied Mathematics CBMS-NSF Monographs.)

Firmani, C., Ghisellini, G., Avila-Reese, V., & Ghirlanda, G. 2006, MNRAS, 370, 185

Flores, H., et al. 2010, GRB Coordinates Network, 11317, 1

Foley, R. J., Chen, H.-W., Bloom, J., & Prochaska, J. X. 2005, GRB Coordinates Network, 3483, 1

Friedman, A. S., & Bloom, J. S. 2005, ApJ, 627, 1

Fugazza, D., et al. 2005, GRB Coordinates Network, 3948, 1

—. 2006, GRB Coordinates Network, 5513, 1

Fynbo, J., Quirion, P.-O., Malesani, D., Thoene, C. C., Hjorth, J., Milvang-Jensen, B., & Jakobsson, P. 2008a, GRB Coordinates Network, 7797, 1

Fynbo, J. P. U., Hjorth, J., Jensen, B. L., Jakobsson, P., Moller, P., & Naranen, J. 2005a, GRB Coordinates Network, 3136, 1

Fynbo, J. P. U., Malesani, D., Hjorth, J., Sollerman, J., & Thoene, C. C. 2008b, GRB Coordinates Network, 8254, 1

Fynbo, J. P. U., Malesani, D., Jakobsson, P., & D'Elia, V. 2009a, GRB Coordinates Network, 9947, 1

Fynbo, J. P. U., Malesani, D., & Milvang-Jensen, B. 2008c, GRB Coordinates Network, 7949, 1

Fynbo, J. P. U., Malesani, D., Thoene, C. C., Vreeswijk, P. M., Hjorth, J., & Henriksen, C. 2006a, GRB Coordinates Network, 5809, 1

Fynbo, J. P. U., et al. 2005b, GRB Coordinates Network, 3749, 1

—. 2006b, A&A, 451, L47

—. 2009b, ApJS, 185, 526

Gehrels, N., et al. 2004, ApJ, 611, 1005

Ghirlanda, G., Ghisellini, G., & Lazzati, D. 2004, ApJ, 616, 331

Goldoni, P., Flores, H., Malesani, D., Levan, A. J., Tanvir, N. R., & Fynbo, J. P. U. 2010, GRB Coordinates Network, 10684, 1

Graham, J. F., Fruchter, A. S., Levan, A. J., Nysewander, M., Tanvir, N. R., Dahlen, T., Bersier, D., & Pe'Er, A. 2007, GRB Coordinates Network, 6836, 1

Grazian, A., et al. 2006, GRB Coordinates Network, 4545, 1

Greiner, J., et al. 2009a, ApJ, 693, 1912

—. 2009b, ApJ, 693, 1610

Grupe, D., Nousek, J. A., vanden Berk, D. E., Roming, P. W. A., Burrows, D. N., Godet, O., Osborne, J., & Gehrels, N. 2007, AJ, 133, 2216

Hughes, G. 1968, Information Theory, IEEE Transactions on, 14, 55

Jakobsson, P., de Ugarte Postigo, A., Gorosabel, J., Tanvir, N., Christensen, L., & Fynbo, J. P. U. 2009, GRB Coordinates Network, 9797, 1

Jakobsson, P., Fynbo, J. P. U., Malesani, D., Hjorth, J., & Milvang-Jensen, B. 2008a, GRB Coordinates Network, 7757, 1

Jakobsson, P., Fynbo, J. P. U., Malesani, D., Tanvir, N. R., Milvang-Jensen, B., Jaunsen, A. O., Vreeswijk, P. M., & Hjorth, J. 2007a, GRB Coordinates Network, 7117, 1

Jakobsson, P., Fynbo, J. P. U., Vreeswijk, P. M., & de Ugarte Postigo, A. 2008b, GRB Coordinates Network, 8077, 1

Jakobsson, P., Levan, A., Chapman, R., Rol, E., Tanvir, N., Vreeswijk, P., & Watson, D. 2006a, GRB Coordinates Network, 5617, 1

Jakobsson, P., Malesani, D., Thoene, C. C., Fynbo, J. P. U., Hjorth, J., Jaunsen, A. O., Andersen, M. I., & Vreeswijk, P. M. 2007b, GRB Coordinates Network, 6283, 1

Jakobsson, P., Vreeswijk, P. M., de Ugarte Postigo, A., Malesani, D., Fynbo, J. P. U., & Sollerman, J. 2008c, GRB Coordinates Network, 8062, 1

Jakobsson, P., Vreeswijk, P. M., Malesani, D., Jaunsen, A. O., Fynbo, J. P. U., Hjorth, J., & Tanvir, N. R. 2008d, GRB Coordinates Network, 7286, 1

Jakobsson, P., Vreeswijk, P. M., Xu, D., & Thoene, C. C. 2008e, GRB Coordinates Network, 7832, 1

Jakobsson, P., et al. 2006b, A&A, 447, 897

—. 2006c, A&A, 460, L13

—. 2007c, GRB Coordinates Network, 6398, 1

—. 2008f, GRB Coordinates Network, 7998, 1

Jaunsen, A. O., Fynbo, J. P. U., Andersen, M. I., & Vreeswijk, P. 2007a, GRB Coordinates Network, 6216, 1

Jaunsen, A. O., Malesani, D., Fynbo, J. P. U., Sollerman, J., & Vreeswijk, P. M. 2007b, GRB Coordinates Network, 6010, 1

Jaunsen, A. O., et al. 2008, ApJ, 681, 453

Kawai, N. 2008, in Astronomical Society of the Pacific Conference Series, Vol. 399, Panoramic Views of Galaxy Formation and Evolution, ed. T. Kodama, T. Yamada, & K. Aoki, 37–+

Kawai, N., Yamada, T., Kosugi, G., Hattori, T., & Aoki, K. 2005, GRB Coordinates Network, 3937, 1

Kelson, D., & Berger, E. 2005, GRB Coordinates Network, 3101, 1

Kocevski, D., & Petrosian, V. 2011, ArXiv e-prints 1110.6175

Koen, C. 2009, MNRAS, 396, 1499

—. 2010, MNRAS, 401, 1369

Kohavi, R. 1995, in International joint Conference on artificial intelligence, Vol. 14, Citeseer, 1137–1145

Kuin, N. P. M., et al. 2009, MNRAS, 395, L21

Lamb, D. Q., & Reichart, D. E. 2000, ApJ, 536, 1

Ledoux, C., Jakobsson, P., Jaunsen, A. O., Thoene, C. C., Vreeswijk, P. M., Malesani, D., Fynbo, J. P. U., & Hjorth, J. 2007, GRB Coordinates Network, 7023, 1

Ledoux, C., Vreeswijk, P., Smette, A., Jaunsen, A., & Kaufer, A. 2006, GRB Coordinates Network, 5237, 1

Ledoux, C., et al. 2005, GRB Coordinates Network, 3860, 1

Levan, A. J., Fynbo, J. P. U., Hjorth, J., Malesani, D., D'Avanzo, P., & D'Elia, V. 2009a, GRB Coordinates Network, 9958, 1

Levan, A. J., Wiersema, K., & Tanvir, N. R. 2011a, GRB Coordinates Network, 12429, 1

Levan, A. J., et al. 2009b, GRB Coordinates Network, 9409, 1

—. 2011b, Science, 333, 199

Levesque, E. M., et al. 2010, MNRAS, 401, 963

Malesani, D., Fynbo, J. P. U., Christensen, L., de Ugarte Postigo, A., Milvang-Jensen, B., & Covino, S. 2009a, GRB Coordinates Network, 9761, 1

Malesani, D., Fynbo, J. P. U., D'Elia, V., de Ugarte Postigo, A., Jakobsson, P., & Thoene, C. C. 2009b, GRB Coordinates Network, 9457, 1

Malesani, D., Fynbo, J. P. U., Jakobsson, P., Vreeswijk, P. M., & Niemi, S.-M. 2008, GRB Coordinates Network, 7544, 1

Malesani, D., Jakobsson, P., Fynbo, J. P. U., Hjorth, J., & Vreeswijk, P. M. 2007, GRB Coordinates Network, 6651, 1

Malesani, D., et al. 2011, GRB Coordinates Network, 11977, 1

McQuinn, M., Lidz, A., Zaldarriaga, M., Hernquist, L., & Dutta, S. 2008, MNRAS, 388, 1101

Melandri, A., Grazian, A., Guidorzi, C., Monfardini, A., Mundell, C. G., & Gomboc, A. 2006, GRB Coordinates Network, 4539, 1

Milne, P. A., & Cenko, S. B. 2011, GRB Coordinates Network, 11708, 1

Milvang-Jensen, B., et al. 2010, GRB Coordinates Network, 10876, 1

Mirabal, N., & Halpern, J. P. 2006, GRB Coordinates Network, 4792, 1

Mirabal, N., Halpern, J. P., & O'Brien, P. T. 2007, ApJ, 661, L127

Miralda-Escude, J. 1998, ApJ, 501, 15

Mortlock, D. J., et al. 2011, Nature, 474, 616

Murakami, T., Yonetoku, D., Izawa, H., & Ioka, K. 2003, PASJ, 55, L65

Oates, S. R., et al. 2009, MNRAS, 395, 490

O'Keefe, P., Gowanlock, M., McConnell, S., & Patton, D. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 411, 318

O'Meara, J., Chen, H.-W., & Prochaska, J. X. 2010, GRB Coordinates Network, 11089, 1

Osip, D., Chen, H.-W., & Prochaska, J. X. 2006, GRB Coordinates Network, 5715, 1

Perley, D. A., Chornock, R., & Bloom, J. S. 2008a, GRB Coordinates Network, 7962, 1

Perley, D. A., Foley, R. J., Bloom, J. S., & Butler, N. R. 2006, GRB Coordinates Network, 5387, 1

Perley, D. A., Prochaska, J. X., Kalas, P., Howard, A., Fitzgerald, M., Marcy, G., & Graham, J. 2009a, GRB Coordinates Network, 10272, 1

Perley, D. A., et al. 2008b, ApJ, 672, 449

—. 2009b, AJ, 138, 1690

—. 2010, MNRAS, 406, 2473

Piranomonte, S., Covino, S., Malesani, D., Fiore, F., Tagliaferri, G., Chincarini, G., & Stella, L. 2006a, GRB Coordinates Network, 5626, 1

Piranomonte, S., Vergani, S. D., Malesani, D., Fynbo, J. P. U., Wiersema, K., & Kaper, L. 2011, GRB Coordinates Network, 12164, 1

Piranomonte, S., et al. 2006b, GRB Coordinates Network, 4520, 1

Price, P. A. 2006, GRB Coordinates Network, 5104, 1

Price, P. A., Berger, E., & Fox, D. B. 2006a, GRB Coordinates Network, 5275, 1

Price, P. A., Cowie, L. L., Minezaki, T., Schmidt, B. P., Songaila, A., & Yoshii, Y. 2006b, ApJ, 645, 851

Prochaska, J. X., Foley, R., Tran, H., Bloom, J. S., & Chen, H.-W. 2006, GRB Coordinates Network, 4593, 1

Prochaska, J. X., Foley, R. J., Holden, B., Magee, D., Cooper, M., & Dutton, A. 2008a, GRB Coordinates Network, 7397, 1

Prochaska, J. X., Perley, D., Howard, A., Chen, H.-W., Marcy, G., Fischer, D., & Wilburn, C. 2008b, GRB Coordinates Network, 8083, 1

Prochaska, J. X., Thoene, C. C., Malesani, D., Fynbo, J. P. U., & Vreeswijk, P. M. 2007, GRB Coordinates Network, 6698, 1

Prochaska, J. X., et al. 2009, ApJ, 691, L27

Quimby, R., Fox, D., Hoeflich, P., Roman, B., & Wheeler, J. C. 2005, GRB Coordinates Network, 4221, 1

Rau, A., Fynbo, J., & Greiner, J. 2010, GRB Coordinates Network, 10350, 1

Richards, J. W., et al. 2011, ApJ, 733, 10

Roming, P. W. A., et al. 2005, Space Sci. Rev., 120, 95

Ruiz-Velasco, A. E., et al. 2007, ApJ, 669, 1

Salvaterra, R., Campana, S., Chincarini, G., Tagliaferri, G., & Covino, S. 2007, MNRAS, 380, L45

Salvaterra, R., et al. 2009, Nature, 461, 1258

Savaglio, S., Palazzi, E., Ferrero, P., & Klose, S. 2007, GRB Coordinates Network, 6166, 1

Schady, P., & Moretti, A. 2006, GRB Coordinates Network, 5296, 1

Schaefer, B. E. 2007, ApJ, 660, 16

Shahmoradi, A., & Nemiroff, R. J. 2011, MNRAS, 411, 1843

Starling, R., Thoene, C. C., Fynbo, J. P. U., Vreeswijk, P., & Hjorth, J. 2006, GRB Coordinates Network, 5131, 1

Stekhoven, D. J., & Bühlmann, P. 2011, ArXiv e-prints 1105.0828

Tanvir, N. R., & Jakobsson, P. 2007, Royal Society of London Philosophical Transactions Series A, 365, 1377

Tanvir, N. R., Wiersema, K., Fynbo, J. P. U., Levan, A. J., & Perley, D. 2010a, GRB Coordinates Network, 11116, 1

Tanvir, N. R., Wiersema, K., & Levan, A. J. 2010b, GRB Coordinates Network, 11230, 1

Tanvir, N. R., Wiersema, K., Levan, A. J., Cenko, S. B., & Geballe, T. 2011, GRB Coordinates Network, 12225, 1

Tanvir, N. R., et al. 2009, Nature, 461, 1254

—. 2010c, GRB Coordinates Network, 11123, 1

Thoene, C. C., de Ugarte Postigo, A., Vreeswijk, P. M., Malesani, D., & Jakobsson, P. 2008a, GRB Coordinates Network, 8058, 1

Thoene, C. C., Fynbo, J. P. U., Jakobsson, P., Vreeswijk, P. M., & Hjorth, J. 2006a, GRB Coordinates Network, 5812, 1

Thoene, C. C., Jakobsson, P., Fynbo, J. P. U., Malesani, D., Hjorth, J., & Vreeswijk, P. M. 2007a, GRB Coordinates Network, 6499, 1

Thoene, C. C., Jaunsen, A. O., Fynbo, J. P. U., Jakobsson, P., & Vreeswijk, P. M. 2007b, GRB Coordinates Network, 6379, 1

Thoene, C. C., Malesani, D., Vreeswijk, P. M., Fynbo, J. P. U., Jakobsson, P., Ledoux, C., & Smette, A. 2008b, GRB Coordinates Network, 7602, 1

Thoene, C. C., Perley, D. A., Cooke, J., Bloom, J. S., Chen, H.-W., & Barton, E. 2007c, GRB Coordinates Network, 6741, 1

Thoene, C. C., et al. 2006b, GRB Coordinates Network, 5373, 1

Totani, T., Kawai, N., Kosugi, G., Aoki, K., Yamada, T., Iye, M., Ohta, K., & Hattori, T. 2006, PASJ, 58, 485

Ukwatta, T. N., Sakamoto, T., Dhuga, K. S., Parke, W. C., Barthelmy, S. D., Gehrels, N., Stamatikos, M., & Tueller, J. 2009, in American Institute of Physics Conference Series, Vol. 1133, American Institute of Physics Conference Series, ed. C. Meegan, C. Kouveliotou, & N. Gehrels, 437–439

Ukwatta, T. N., et al. 2010, ApJ, 711, 1073

—. 2011, MNRAS, 1618

vanden Berk, D. E., Grupe, D., Racusin, J., Roming, P., & Koch, S. 2008, in American Institute of Physics Conference Series, Vol. 1000, American Institute of Physics Conference Series, ed. M. Galassi, D. Palmer, & E. Fenimore, 80–83

Vergani, S. D., et al. 2010, GRB Coordinates Network, 10495, 1

Vreeswijk, P., Jakobsson, P., Ledoux, C., Thoene, C., & Fynbo, J. 2006, GRB Coordinates Network, 5535, 1

Vreeswijk, P., Malesani, D., Fynbo, J., Jakobsson, P., Thoene, C., Sollerman, J., Watson, D., & Milvang-Jensen, B. 2008a, GRB Coordinates Network, 8301, 1

Vreeswijk, P. M., Fynbo, J. P. U., Malesani, D., Hjorth, J., & de Ugarte Postigo, A. 2008b, GRB Coordinates Network, 8191, 1

Vreeswijk, P. M., Milvang-Jensen, B., Smette, A., Malesani, D., Fynbo, J. P. U., Jakobsson, P., Jaunsen, A. O., & Ledoux, C. 2008c, GRB Coordinates Network, 7451, 1

Vreeswijk, P. M., Thoene, C. C., Malesani, D., Fynbo, J. P. U., Hjorth, J., Jakobsson, P., Tanvir, N. R., & Levan, A. J. 2008d, GRB Coordinates Network, 7601, 1

Vreeswijk, P. M., et al. 2007, A&A, 468, 83

—. 2008e, GRB Coordinates Network, 7444, 1

Wiersema, K., Graham, J., Tanvir, N., Rol, E., Levan, A., & Fruchter, A. 2008a, GRB Coordinates Network, 7818, 1

Wiersema, K., Levan, A., Kamble, A., Tanvir, N., & Malesani, D. 2009, GRB Coordinates Network, 9673, 1

Wiersema, K., Tanvir, N., Vreeswijk, P., Fynbo, J., Starling, R., Rol, E., & Jakobsson, P. 2008b, GRB Coordinates Network, 7517, 1

Wiersema, K., et al. 2011, GRB Coordinates Network, 12431, 1

Wijers, R. A. M. J., Bloom, J. S., Bagla, J. S., & Natarajan, P. 1998, MNRAS, 294, L13

Xiao, L., & Schaefer, B. E. 2009, ApJ, 707, 387

—. 2011, ApJ, 731, 103

Xu, D., et al. 2009, GRB Coordinates Network, 10053, 1

Zhang, Z., Deng, J., Lu, R., & Gao, H. 2006, Chinese J. Astron. Astrophys., 6, 312

Table 6.   Validation Redshifts and Predictions

GRB	$\widehat{\mathcal{Q}}$	z	References
100728B	5.63e-01	2.106	Flores et al. 2010
100814A	7.04e-01	1.44	O'Meara et al. 2010
100816A	9.33e-01	0.8035	Tanvir et al. 2010a,c
100901A	4.22e-01	1.408	Chornock et al. 2010a
100906A	9.19e-01	1.727	Tanvir et al. 2010b
101219B	6.07e-01	0.5519	de Ugarte Postigo et al. 2011c
110205A	3.19e-01	2.22	Cenko et al. 2011
110213A	8.89e-01	1.46	Milne & Cenko 2011
110422A	9.33e-01	1.770	Malesani et al. 2011; de Ugarte Postigo et al. 2011a
110503A	9.33e-01	1.613	de Ugarte Postigo et al. 2011b
110715A	8.89e-01	0.82	Piranomonte et al. 2011
110726A	5.56e-01	1.036	Cucchiara et al. 2011a
110731A	9.33e-01	2.83	Tanvir et al. 2011
110801A	8.67e-01	1.858	Cabrera Lavers et al. 2011
110808A	5.63e-01	1.348	de Ugarte Postigo et al. 2011d